

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:52:32 ; Search time 47 Seconds
(without alignments)
2559.888 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLKPIVLSILLINTEPL.....FLSDTPQMGSRFTGGVNVK 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4036	100.0	758	21	AA18719
2	3845	95.3	764	23	AA18719
3	3836	95.0	758	23	AAU73011
4	3819	94.6	758	21	AA174566
5	3816	94.5	764	21	AA174557
6	3814	94.5	758	21	AA174565
7	3751	92.9	758	21	AA174564
8	3714	92.0	735	21	AA174556
9	3644	90.3	728	24	ABP80923

10	1808	44.8	339	21	AA174562	Neisseria meningit
11	1791	44.4	339	21	AA174563	Neisseria meningit
12	1763	43.7	339	21	AA174561	Neisseria gonorrhoe
13	1580.5	39.2	814	23	AA017586	M catarrhalis MCAI
14	1580.5	39.2	818	22	AA119945	Moraxella catarrha
15	1202	29.8	913	22	AA119946	Moraxella catarrha
16	1202	29.8	913	23	AA017585	M catarrhalis MCAI
17	763	18.9	174	21	AA174555	Neisseria gonorrhoe
18	216	5.4	713	17	AAW05181	Neisseria gonorrhoe
19	216	5.4	713	22	AAE05408	Neisseria gonorrhoe
20	216	5.4	713	24	ABP77241	N. gonorrhoeae or N
21	214.5	5.3	720	23	ABP78069	N. gonorrhoeae am
22	214	5.3	719	17	AAW01462	Amino acid sequenc
23	212.5	5.3	703	21	AA175567	NTHI HxuC protein.
24	212.5	5.3	703	22	AA175567	Neisseria meningit
25	212.5	5.3	714	23	AAU73008	Neisseria meningit
26	208.5	5.2	703	21	AA175568	Neisseria meningit
27	208.5	5.2	703	22	AA175568	Neisseria meningit
28	206	5.1	702	22	AA175568	Neisseria meningit
29	205.5	5.1	703	22	AA175568	Neisseria meningit
30	205	5.1	702	22	AA175568	Neisseria meningit
31	203	5.0	702	22	AA175568	Neisseria meningit
32	203	5.0	702	22	AA175568	Neisseria meningit
33	203	5.0	725	17	AAW01460	Neisseria meningit
34	202	5.0	713	17	AAW05180	Hib HxuC protein.
35	202	5.0	713	22	AAE05407	Neisseria gonorrhoe
36	202	5.0	713	23	ABG91066	N. gonorrhoeae or N
37	200.5	5.0	703	22	AA175568	Neisseria gonorrhoe
38	200.5	5.0	725	20	AA175568	Neisseria meningit
39	198.5	4.9	687	22	AAE05851	Neisseria meningit
40	198.5	4.9	708	20	AA174501	Pseudomonas stutze
41	198.5	4.9	772	20	AA174374	Pseudomonas stutze
42	194	4.8	757	20	AA174472	Pseudomonas stutze
43	194	4.8	763	20	AA174472	Pseudomonas stutze
44	189	4.7	715	17	AAW01461	NTHI HxuC protein.
45	189	4.7	725	20	AA174472	Neisseria meningit

ALIGNMENTS

RESULT 1

AA18719	AA18719 standard; Protein; 758 AA.
ID	AA18719 standard; Protein; 758 AA.
XX	
AC	AA18719; (first entry)
XX	
DT	22-JAN-2001 (first entry)
XX	
DE	A Neisseria meningitidis BASB082 polypeptide.
XX	
KW	BASB082; BASB083; BASB091; BASB092; BASB0101; infection; vaccine; gene therapy; upper respiratory tract infection; bacteremia; meningitis; invasive bacterial disease.
XX	
OS	Neisseria meningitidis.
XX	
PN	WO200055327-A2.
XX	
PD	21-SEP-2000.
XX	
PF	07-MAR-2000; 2000WO-EP01955.
XX	
PR	12-MAR-1999; 99GB-0005815.
PR	21-APR-1999; 99GB-0009094.
PR	23-APR-1999; 99GB-0009503.
PR	28-APR-1999; 99GB-0009787.
XX	
XX	07-MAY-1999; 99GB-0010710.
PA	(SMIX) SMITHKLINE BEECHAM BIOLOGICALS.
XX	
PI	Defrenne C, Delmelle C, Ruelle J;
XX	

DR WPI; 2000-602119/57.
 DR N-PSDB; AAA75744.
 XX Novel polypeptides designated BASB 082, 083, 091, 092, and 101 derived
 PT from meningococcus bacterium useful for producing vaccines against
 PT infections and in diagnostic assays -
 XX Claim 3; Page 99-101; 108pp; English.
 XX The present sequence represents a BASB082 polypeptide. The specification
 CC describes BASB082, BASB083, BASB091, BASB092, and BASB101 polypeptides.
 CC The polynucleotides and polypeptides are useful as diagnostic reagents.
 CC and for diagnosing N. meningitidis infection. The polynucleotides may
 CC be used as hybridisation probe for RNA, cDNA and genomic DNA to isolate
 CC full-length cDNAs and genomic clones encoding BASB082, BASB083, BASB091,
 CC BASB092 or BASB101 polypeptides and to isolate cDNA and genomic clones
 CC of other genes that have a high identity particularly high sequence
 CC identity to BASB082, BASB083, BASB091, BASB092 or BASB101 genes. The
 CC vaccine compositions are useful for inducing an immunological response
 CC in humans. The polynucleotides encoding BASB082, BASB083, BASB091,
 CC BASB092 or BASB101 polypeptides are useful in gene therapy to induce
 CC an immunological response. The polypeptides are useful for treating
 CC upper respiratory tract infection, invasive bacterial diseases, such as
 CC bacteremia and meningitis.
 XX Sequence 758 AA;
 SQ Query Match 100.0%; Score 4036; DB 21; Length 758;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQTLKPIVLSTILLINTPLLAQAHETEQSVGLTETVWVKSRPRATSGLLHTSTASDKI 60
 DB 1 MAQTLKPIVLSTILLINTPLLAQAHETEQSVGLTETVWVKSRPRATSGLLHTSTASDKI 60

QY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIKVLNHHGETGMADF 120
 DB 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIKVLNHHGETGMADF 120

QY 121 SPDHAIWVDTALSQVEILRGPVTLTSSGNVAGLVDVADGKIPKMPENGVSGLGLRL 180
 DB 121 SPDHAIWVDTALSQVEILRGPVTLTSSGNVAGLVDVADGKIPKMPENGVSGLGLRL 180

QY 181 SSGNLEKLTSGGINIGLGNFVLTGELYKSGDYAVPRYNLKRLPDSPRPFANGOHRA 240
 DB 181 SSGNLEKLTSGGINIGLGNFVLTGELYKSGDYAVPRYNLKRLPDSPRPFANGOHRA 240

QY 241 VLGRKKEFYERTYSDDRDQYGLPAHSHEYDDCHADIIWQSLINKRYLQLYPHLLTDEVD 300
 DB 241 VLGRKKEFYERTYSDDRDQYGLPAHSHEYDDCHADIIWQSLINKRYLQLYPHLLTDEVD 300

QY 301 DYDNPGLSCGFHDDDDAHAAHNGKPIDLNRKRYELRAEWKQFPFGFEALRVHLNRNDY 360
 DB 301 DYDNPGLSCGFHDDDDAHAAHNGKPIDLNRKRYELRAEWKQFPFGFEALRVHLNRNDY 360

QY 361 HHDEKAGDAVENFNNQTONARIELRHQPIGRKAGSGVQYLGOKSSALSATSVAVKQPM 420
 DB 361 HHDEKAGDAVENFNNQTONARIELRHQPIGRKAGSGVQYLGOKSSALSATSVAVKQPM 420

QY 421 LLDNKVQHYFFGVEQANWDFNFTLEGVVRKQKASIRYDKALIDRNYKQPLPDILGAH 480
 DB 421 LLDNKVQHYFFGVEQANWDFNFTLEGVVRKQKASIRYDKALIDRNYKQPLPDILGAH 480

QY 481 ROTARSAFALSGNWFYFTQHKLSLTASHQERLPSTQELIYAHGKHVATNTFEVGNKHLNKR 540
 DB 481 ROTARSAFALSGNWFYFTQHKLSLTASHQERLPSTQELIYAHGKHVATNTFEVGNKHLNKR 540

QY 541 SNNIELALGVEGDRWQYNLALYRNFNGYIYAQTLNDGRGPKSTEDDSEMKLVRYNOSGA 600
 DB 541 SNNIELALGVEGDRWQYNLALYRNFNGYIYAQTLNDGRGPKSTEDDSEMKLVRYNOSGA 600

QY 601 DFYGAEGEYIFKPTPRYRIGVSGDYVRGRLNKLPFLGREDAYGNRPFFIAQDDQNAVRVP 660
 DB 601 DFYGAEGEYIFKPTPRYRIGVSGDYVRGRLNKLPFLGREDAYGNRPFFIAQDDQNAVRVP 660

DB 601 DFYGAEGEYIFKPTPRYRIGVSGDYVRGRLNKLPFLGREDAYGNRPFFIAQDDQNAVRVP 660
 QY 661 AARLGFHLKASLTDRIIDANLDYRVRFAQNKLAARYETPTPGHMLNLGANYRRNTRYGEWN 720
 DB 661 AARLGFHLKASLTDRIIDANLDYRVRFAQNKLAARYETPTPGHMLNLGANYRRNTRYGEWN 720
 QY 721 WYKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
 DB 721 WYKADNLLNQSVYAHSSFLSDTPQMGRSFTGGVNVKF 758

RESULT 2
 ABB78068
 ID ABB78068 standard; Protein; 764 AA.
 XX ABB78068;
 XX 05-NOV-2002 (first entry)
 XX Amino acid sequence of p88 polypeptide.
 DE
 XX p177; p88; p64; p55; p46; vaccine; gonorrhea.
 OS Neisseria gonorrhoeae.
 XX WO200260936-A2.
 XX 08-AUG-2002.
 XX 31-JAN-2002; 2002WO-USO2881.
 XX 31-JAN-2001; 2001US-266070P.
 PR 06-AUG-2001; 2001US-310356P.
 PR 23-OCT-2001; 2001US-344452P.
 XX (TOWA) UNIV IOWA RES FOUND.
 PA (REGC) UNIV CALIFORNIA.
 PA (APIC) APICELLA M A.
 PA (EDWA) EDWARDS J L.
 PA (GIBS) GIBSON B W.
 PA (SCHE) SCHEFFLER K.
 PA (BROW) BROWN E.
 XX Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;
 PI WPI; 2002-619227/66.
 DR N-PSDB; ABQ78299.
 DR New polypeptide comprising p177, p88, p64, p55 or p46 from Neisseria
 PT gonorrhoeae, useful for preventing, or protecting a female patient
 PT against, N. gonorrhoeae colonization or infection -
 XX Claim 7; Page 115-117; 130pp; English.
 CC The present sequence represents a p88 polypeptide. The specification
 CC describes p177, p88, p64, p55 and p46 polypeptides from Neisseria
 CC gonorrhoeae. The polypeptides are useful as vaccines, for preventing,
 CC or protecting a female patient against, N. gonorrhoeae colonization or
 CC infection. Such immunisation can prevent gonorrhea in women.
 SQ Sequence 764 AA;

Query Match 95.3%; Score 3845; DB 23; Length 764;
 Best Local Similarity 95.9%; Pred. No. 0;
 Matches 730; Conservative 7; Mismatches 18; Indels 6; Gaps 3;

QY 1 MAQTLKPIVLSTILLINTPLLAQAHETEQSVGLTETVWVKSRPRATSGLLHTSTASDKI 60
 DB 7 MAQTLKPIVLSTILLINTPLLSQAHCETQSVGLTETVWVKSRPRATSGLLHTSTASDKI 66

QY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIKVLNHHGETGMADF 120
 DB 67 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIKVLNHHGETGMADF 126

Db 658 RVPAAALGFHFKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNLGANYRNRTRYG 717
 QY 718 EWNWYVKADNLLNQSVYAHSSFLSDTTPQMGRSFTGGVNVKF 758
 Db 718 EWNWYVKADNLLNQSVYAHSSFLSDTTPQMGRSFTGGVNVKF 758
 RESULT 4
 AAY74566
 ID AAY74566 standard; Protein; 758 AA.
 XX AC AAY74566;
 XX DE 21-MAR-2000 (first entry)
 XX DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:606.
 XX DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX OS Neisseria meningitidis.
 XX PN WO9957280-A2.
 XX PD 11-NOV-1999.
 XX PF 30-APR-1999; 99W0-US09346.
 XX PR 01-MAY-1998; 98W0-0083758.
 XX PR 31-JUL-1998; 98US-0094869.
 XX PR 02-SEP-1998; 98US-0098994.
 XX PR 02-SEP-1998; 98US-0099062.
 XX PR 09-OCT-1998; 98US-0103749.
 XX PR 09-OCT-1998; 98US-0103794.
 XX PR 09-OCT-1998; 98US-0103796.
 XX PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 XX (GENO-) INST GENOMIC RES.
 XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 XX DR N-PSDE; AAZ53328.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX Claim 2; Page 424-425; 1453pp; English.
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX SQ Sequence 758 AA;
 Query Match 94.6%; Score 3819; DB 21; Length 758;
 Best Local Similarity 95.5%; Pred.No.0;
 Matches 72; Conservative 7; Mismatches 21; Indels 6; Gaps 3;

QY 1 MAQTTLPKPIVLSILLINTPLLAQAHTEQSGLTETVTWVSKSRPRATSGLLHTSTASDKI 60
 Db 1 MAQTTLPKPIVLSILLINTPLLSQAHGTEQSGLTETVSVGKSRPRATSGLLHTSTASDKI 60
 QY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGQTGRRIKVLNHHGETGDMADF 120
 Db 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGQTGRRIKVLNHHGETGDMADF 120
 QY 121 SPDHAIWVDTALSOQVEILRGPVTLTYSSGNVAGLVADVADGKIPEKMPENCVSGELGURL 180
 Db 121 SPDHAIWVDTALSOQVEILRGPVTLTYSSGNVAGLVADVADGKIPEKMPENCVSGELGURL 180
 QY 181 SSGNLEKLTSGGINIGLGNFVLTHTGLYRKSGDYAVPRYRNLRKLPDPSPRFANGQHRA 240
 Db 181 SSGNLEKLTSGGINIGLGNFVLTHTGLYRKSGDYAVPRYRNLRKLPDPS--HADSTQTS 237
 QY 241 V-LGW--RKFYRRTYSDRRDOVGLPAHSHEYDDCHADIWOKSLINKRYLQLYPHLLTE 297
 Db 238 IGLSWGVEKFGIGAAYSDRRDOVGLPAHSHEYDDCHADIWOKSLINKRYLQLYPHLLTE 297
 QY 298 EDVDYDNPGLSCGPHDDDDAHAAHNGKPMWDLRNKRYELRAEWKQPPGFEALRVHLNR 357
 Db 298 EDIDYDNPGLSCGPHDDDDAHAAHNGKPMWDLRNKRYELRAEWKQPPGFEALRVHLNR 357
 QY 358 NDYHDEKAGDAVENFFNNTQONARIELRHQPIGRKLGSGWGVQYLGQKSSALSATSEAVK 417
 Db 358 NDYHDEKAGDAVENFFNNTQONARIELRHQPIGRKLGSGWGVQYLGQKSSALSATSEAVK 417
 QY 418 QPMLLDNKVQHYSPFGEQANWNTLEGGVVRVEKOKASTRYDKALIDRENNYKQPLPDL 477
 Db 418 QPMLLDNKVQHYSPFGEQANWNTLEGGVVRVEKOKASTRYDKALIDRENNYKQPLPDL 477
 QY 478 GAHQRTARSFALSGNWFYFTQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVGNKHLN 537
 Db 478 GAHQRTARSFALSGNWFYFTQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVGNKHLN 537
 QY 538 KERSNNIELALGYEGDRWQYNLALYRNRFNGYIYAQTLNDGRGPKSIEDDSEMKLVRYNQ 597
 Db 538 KERSNNIELALGYEGDRWQYNLALYRNRFNGYIYAQTLNDGRGPKSIEDDSEMKLVRYNQ 597
 QY 598 SGADFYGAEGEYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPPIAQQDNAP 657
 Db 598 SGADFYGAEGEYFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPPIAQQDNAP 657
 QY 658 RVPAAALGFHFKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNLGANYRNRTRYG 717
 Db 658 RVPAAALGFHFKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNLGANYRNRTRYG 717
 QY 718 EWNWYVKADNLLNQSVYAHSSFLSDTTPQMGRSFTGGVNVKF 758
 Db 718 EWNWYVKADNLLNQSVYAHSSFLSDTTPQMGRSFTGGVNVKF 758
 RESULT 5
 AAY74557
 ID AAY74557 standard; Protein; 764 AA.
 XX AC AAY74557;
 XX DT 21-MAR-2000 (first entry)
 XX DE Neisseria meningitidis ORF 147 protein sequence SEQ ID NO:588.
 XX DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX OS Neisseria meningitidis.
 XX PN WO9957280-A2.
 XX PD 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR N-PSDB; AAZ53319.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX Claim 2; Page 414-415; 1453pp; English.
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
 CC represent novel Neisseria meningitidis and N. gonorrhoea polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ54615 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX Sequence 764 AA;
 SQ
 Query March 94.5%; Score 3816; DB 21; Length 764;
 Best Local Similarity 95.4%; Pred. No. 0;
 Matches 726; Conservative 7; Mismatches 22; Indels 6; Gaps 3;
 QY 1 MAQTLKPIVLSILLINPLLAQAHEHQSVGLTETVTVGKSRPRATSGLLHTSTASDKI 60
 Db 7 MAQTLKPIVLSILLINPLLSQAHEHQSVGLTETVTVGKSRPRATSGLLHTSTASDKI 66
 QY 61 ISGDTLROKAVNLGDALGVPGIHASQVGGASAPVIRGQTRIKVLNHHGETGDMADF 120
 Db 67 ISGDTLROKAVNLGDALGVPGIHASQVGGASAPVIRGQTRIKVLNHHGETGDMADF 126
 QY 121 SPDHAIWVDTALSQVEILRGVETLLYSNGVAGLVADVADGKIPEKMPENGVSCELGLRL 180
 Db 127 SPDHAIWVDSALSQVEILRGVETLLYSNGVAGLVADVADGKIPEKMPENGVSCELGLRL 186
 QY 181 SSGNLEKLTSGINIGLKGNFVILHTEGLYRSGYAVPRYENLKLPSRPFANGOHRA 240
 Db 187 SSGNLEKLTSGINIGLKGNFVILHTEGLYRSGYAVPRYENLKLPSRPFANGOHRA 243
 QY 241 V-LGW--RKRFRVRYTSRRDQYGLPAHSHEYDCHADIIWOKSLINKRYILQLYPHLLITE 297
 Db 244 IGLSWGKGFGIAGYSDRQYGLPAHSHEYDCHADIIWOKSLINKRYILQLYPHLLITE 303
 QY 298 EDVYDNPGLSCGFHDDDAHAHANGKFWIDLRNRYELRAEWKQPPPGFEALRVHLNR 357
 Db 304 EDIDYDNPGLSCGFHDDDAHAHANGKFWIDLRNRYELRAEWKQPPPGFEALRVHLNR 363
 QY 358 NDYHDEKAGDAVENFFNQNTQARIELRHQPIGLKSGWGVQVYLGOKSSALSATSBAVK 417
 Db 364 NDYHDEKAGDAVENFFNQNTQARIELRHQPIGLKSGWGVQVYLGOKSSALSATSBAVK 423

QY 418 QPMLLDNKVQHYSPFGEQANWDFLLEGGVVEKOKASIRYDKALIDRENYKQPLDL 477
 Db 424 QPMLLDNKVQHYSPFGEQANWDFLLEGGVVEKOKASIRYDKALIDRENYKQPLDL 483
 QY 478 GAHQRTARSFALSGNWNFTPOHKLSLTASHOERLPSLOELYAHGHVATNTFEVGNKHLN 537
 Db 484 GAHQRTARSFALSGNWNFTPOHKLSLTASHOERLPSLOELYAHGHVATNTFEVGNKHLN 543
 QY 538 KERSNNIELALGYEGDRWQYNLALYRNFRGNYIYAQTLDGGRPKSIEDDSEMKLVRYNQ 597
 Db 544 KERSNNIELALGYEGDRWQYNLALYRNFRGNYIYAQTLDGGRPKSIEDDSEMKLVRYNQ 603
 QY 598 SGADFYGAEGETYFKPTPRYRIGVSGDYVRGRLKWLPSLPGRDAYGNRPPIAQQONAP 657
 Db 604 SGADFYGAEGETYFKPTPRYRIGVSGDYVRGRLKWLPSLPGRDAYGNRPPIAQQONAP 663
 QY 658 RYPAARLGFHLKASLTDRIDANLDYRVYFAQNKLYRTRTPGHMLNIGANYRENTVYG 717
 Db 664 RYPAARLGVHLKASLTDRIDANLDYRVYFAQNKLYRTRTPGHMLNIGANYRENTVYG 723
 QY 718 ENWVYKADNLLNQSVYAHSSFLSDTPQWGRSFTGGVNVKF 758
 Db 724 RNWVYKADNLLNQSVYAHSSFLSDTPQWGRSETGGVNVKF 764
 RESULT 6
 AAY74565
 ID AAY74565 standard; Protein; 758 AA.
 XX AAY74565;
 AC AAY74565;
 DT 21-MAR-2000 (first entry)
 XX Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:604.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX Neisseria meningitidis.
 OS Neisseria meningitidis.
 XX WO9957280-A2.
 PN 11-NOV-1999.
 XX 30-APR-1999; 99WO-US09346.
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR N-PSDB; AAZ53327.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX Claim 2; Page 422-423; 1453pp; English.
 PS AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC

CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA25437 to AA25476 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 758 AA;
 SQ Query Match 94.5%; Score 3814; DB 21; Length 758;
 Best Local Similarity 95.3%; Pred. No. 0;
 Matches 725; Conservative 10; Mismatches 20; Indels 6; Gaps 3;

QY 1 MAQTLKPVLISILLINTPLLAQAHEHQSVGLTVTVVVGKSRPRATSGLLHTSTASDKI 60
 DB 1 MAQTLKPVLISILLINTPLLAQAHEHQSVGLTVTVVVGKSRPRATSGLLHTSTASDKI 60
 QY 61 ISGDTLRQKAVNLGDALDGVPGIHAHQYGGGASAPVIRGQTGRRIKVLNHHGETGDMADF 120
 DB 61 ISGDTLRQKAVNLGDALDGVPGIHAHQYGGGASAPVIRGQTGRRIKVLNHHGETGDMADF 120
 QY 121 SPDAIMVDTALSQVEILRGVPTLLSYSSGNVAGLVADGKIPEKMPGVSGELGLRL 180
 DB 121 SPDAIMVDTALSQVEILRGVPTLLSYSSGNVAGLVADGKIPEKMPGVSGELGLRL 180
 QY 181 SSGNLEKTSNGINTGLKGNVHTEGLYKSGDVAVPRYLNKRLPDSRPFANGQHRA 240
 DB 181 SSGNLEKTSNGINTGLKGNVHTEGLYKSGDVAVPRYLNKRLPDSRPFANGQHRA 240
 QY 241 V-LGW--RRFRYRTYSRDRDQYGLPAHSHYEDDCHADIIWQSLINKRYLYQLYPHLLTE 297
 DB 238 IGLSVGKEGFGIYVAYSDRDRDQYGLPAHSHYEDDCHADIIWQSLINKRYLYQLYPHLLTE 297
 QY 298 EVDVDNPGSCGFHDDDAHAHANGKFWIDLNKRYELRAEWKQPPFGFALRVLHNR 357
 DB 298 EDIDYDNPGLSCGFHDDDAHAHANGKFWIDLNKRYELRAEWKQPPFGFALRVLHNR 357
 QY 358 NDYHDEKAGDAVENFNNQTNARIELRHQPTIGRLKSGWQYLGOKSSALSATSEAVK 417
 DB 358 NDYHDEKAGDAVENFNNQTNARIELRHQPTIGRLKSGWQYLGOKSSALSATSEAVK 417
 QY 418 QPMLLDNKVQHYSSFFGVEQANNDNFTLEGGRVVEKQASIRYDKALIDRENYKQPLPDL 477
 DB 418 QPMLLDNKVQHYSSFFGVEQANNDNFTLEGGRVVEKQASIRYDKALIDRENYKQPLPDL 477
 QY 478 GAHQRTARSFALSGNWFYTPQHKLSLTASHOBLPSTQELYAHGKHVATNTEVGNKHLN 537
 DB 478 GAHQRTARSFALSGNWFYTPQHKLSLTASHOBLPSTQELYAHGKHVATNTEVGNKHLN 537
 QY 538 KERSNNIELALGVEGRWQYNLALYRNRRGNVYIAQTLNDGRPKSIEDDSEMKLVRYNQ 597
 DB 538 KERSNNIELALGVEGRWQYNLALYRNRRGNVYIAQTLNDGRPKSIEDDSEMKLVRYNQ 597
 QY 598 SGADFYGABGEIYFKPTPRYRGVSDYVRGRILKNLPSLPGREDAYGNRPFFIAQDDQNA 657
 DB 598 SGADFYGABGEIYFKPTPRYRGVSDYVRGRILKNLPSLPGREDAYGNRPFFIAQDDQNA 657
 QY 658 RYPAARLGHFLKASLTDRLDANLDYRVFAQNKLAERYETRTPHHMLNLGANYRNRTRYG 717
 DB 658 RYPAARLGHFLKASLTDRLDANLDYRVFAQNKLAERYETRTPHHMLNLGANYRNRTRYG 717
 QY 718 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
 DB 718 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758

AAAY74564
 ID AAAY74564 standard; Protein; 758 AA.
 XX
 AC AAAY74564;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE *Neisseria gonorrhoeae* ORF 149 protein sequence SEQ ID NO:602.
 XX
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS *Neisseria gonorrhoeae*.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR N-PSDB; AA253326.
 XX
 PT Novel *Neisseria* polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 421-422; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA254941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 758 AA;
 Query Match 92.9%; Score 3751; DB 21; Length 758;
 Best Local Similarity 93.3%; Pred. No. 0;
 Matches 710; Conservative 13; Mismatches 32; Indels 6; Gaps 3;
 QY 1 MAQTLKPVLISILLINTPLLAQAHEHQSVGLTVTVVVGKSRPRATSGLLHTSTASDKI 60
 DB 1 MAQTLKPVLISILLINTPLLAQAHEHQSVGLTVTVVVGKSRPRATSGLLHTSTASDKI 60
 QY 61 ISGDTLRQKAVNLGDALDGVPGIHAHQYGGGASAPVIRGQTGRRIKVLNHHGETGDMADF 120
 DB 61 ISGDTLRQKAVNLGDALDGVPGIHAHQYGGGASAPVIRGQTGRRIKVLNHHGETGDMADF 120

QY 121 SPDHAIMVDTALSSQVEILRGVPTLLYSNGVAGLVADVADGKIPKMPENGVSGBGLRL 180
 Db 121 SPDHAIMVDTALSSQVEILRGVPTLLYSNGVAGLVADVADGKIPKMPENGVSGBGLRL 180
 QY 181 SSGNLEKLTSGGINIGLGNFVHLTEGLYRKSGDYAVPRYRLKELPDSPPRFANGQRA 240
 Db 181 SSGNLEKLTSGGINIGLGNFVHLTEGLYRKSGDYAVPRYRLKELPDSPPRFANGQRA 240
 QY 241 V-LGW--RRFYRRRTYSDRRDQVGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTE 297
 Db 241 V-LGW--RRFYRRRTYSDRRDQVGLPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTE 297
 QY 298 EDVDYDNPGLSCGFHDDDAHAHANGKPMWIDLRNRYELRAEWQPPFGFPAALRVHLNR 357
 Db 298 EDVDYDNPGLSCGFHDDDAHAHANGKPMWIDLRNRYELRAEWQPPFGFPAALRVHLNR 357
 QY 358 NDYHDEKAGDAVENFNQOTVARIELRHQPIGLKSGWGVQYLGOKSSALSATSEAVK 417
 Db 358 NDYHDEKAGDAVENFNQOTVARIELRHQPIGLKSGWGVQYLGOKSSALSATSEAVK 417
 QY 418 QPMLLDNKVQHYSGFVEQANWDFTEGGYRVEKOKASIRYDKALIDRENYKQPLPDL 477
 Db 418 QPMLLDNKVQHYSGFVEQANWDFTEGGYRVEKOKASIRYDKALIDRENYKQPLPDL 477
 QY 478 GAHQRTARSPALSGNMYFTPOHKLSTASHOERLPSTQELYAHGKHVATNTFEVGNKHLN 537
 Db 478 GAHQRTARSPALSGNMYFTPOHKLSTASHOERLPSTQELYAHGKHVATNTFEVGNKHLN 537
 QY 538 KERSNNIELALGYEGDRWQVNLALYRNFNGYIYAQTLDNGRGPKSIEDDSEMKLVRYNQ 597
 Db 538 KERSNNIELALGYEGDRWQVNLALYRNFNGYIYAQTLDNGRGPKSIEDDSEMKLVRYNQ 597
 QY 598 SGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGRDAGNRPFTAQDQNP 657
 Db 598 SGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGRDAGNRPFTAQDQNP 657
 QY 658 RVPARLGFHLKASLTDRIANDLDYRFAQNKLAETRTPEGHMLNLGANYRNRTRYG 717
 Db 658 RVPARLGFHLKASLTDRIANDLDYRFAQNKLAETRTPEGHMLNLGANYRNRTRYG 717
 QY 718 EWNWYVKADNLNQSVYAHSSFLSDTPQMGSRSTGGVNVKF 758
 Db 718 EWNWYVKADNLNQSVYAHSSFLSDTPQMGSRSTGGVNVKF 758

RESULT 8
 AAY74556
 ID AAY74556 standard; Protein; 735 AA.
 AC AAY74556;
 DT 21-MAR-2000 (first entry)
 XX Neisseria meningitidis ORF 147 protein sequence SEQ ID NO:586.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX Neisseria meningitidis.
 OS Neisseria meningitidis.
 XX Neisseria meningitidis.
 PN WO9957280-A2.
 PD 11-NOV-1999.
 XX 11-NOV-1999.
 XX 30-APR-1999; 99WO-US09346.
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR N-FSDB; AAZ53318.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX Claim 2; Page 413; 1453pp; English.
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX Sequence 735 AA;
 SQ Query Match 92.0%; Score 3714; DB 21; Length 735;
 Best Local Similarity 95.3%; Pred. No. 0;
 Matches 702; Conservative 10; Mismatches 19; Indels 6; Gaps 3;
 QY 25 HETQSVGLEFVTVVGVKSRPRATSGLLHTSTADSKIIISGDTLRKAVNLGDALDGVGIIH 84
 Db 2 HKTWSVDLETVSVVGVKSRPRATSGLLHTSTADSKIIISGDTLRKAVNLGDALDGVGIIH 61
 QY 85 ASQYGGGASAPVIRGQTRRIKVLNHHGETGDMADFSFDHAIMVDTALSSQVEILRGVPT 144
 Db 62 ASQYGGGASAPVIRGQTRRIKVLNHHGETGDMADFSFDHAIMVDTALSSQVEILRGVPT 121
 QY 145 LLYSSGNVAGLVADVADGKIPKMPENGVSGBGLRLSGNLEKLTSGGINIGLGNFVHL 204
 Db 122 LLYSSGNVAGLVADVADGKIPKMPENGVSGBGLRLSGNLEKLTSGGINIGLGNFVHL 181
 QY 205 TEGLYRKSGDYAVPRYRLKELPDSPPRFANGQRAV-LGW--RRFYRRRTYSDRRDQY 261
 Db 182 TEGLYRKSGDYAVPRYRLKELPDSPPRFANGQRAV-LGW--RRFYRRRTYSDRRDQY 238
 QY 262 LPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHA 321
 Db 239 LPAHSHEDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHA 298
 QY 322 HNGKPMWIDLRNRYELRAEWQPPFGFPAALRVHLNRNDYHDEKAGDAVENFNQOTNA 381
 Db 299 HNGKPMWIDLRNRYELRAEWQPPFGFPAALRVHLNRNDYHDEKAGDAVENFNQOTNA 358
 QY 382 RIELRHQPIGLKSGWGVQYLGOKSSALSATSEAVKQPLMLLDNKVQHYSGFVEQANW 441
 Db 359 RIELRHQPIGLKSGWGVQYLGOKSSALSATSEAVKQPLMLLDNKVQHYSGFVEQANW 418
 QY 442 FTLEGGYRVEKOKASIRYDKALIDRENYKQPLPDLGAHQRTARSPALSGNMYFTPOHKL 501
 Db 419 FTLEGGYRVEKOKASIRYDKALIDRENYKQPLPDLGAHQRTARSPALSGNMYFTPOHKL 478
 QY 502 SLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLNERSNNIELALGYEGDRWQVNLAL 561
 Db 479 SLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLNERSNNIELALGYEGDRWQVNLAL 538

562 YRNRFGNYIAQTLNDGRGPKSIEDDSEMKLVRYNQGADFYGAEGEYIFKPTPRYRIGV 621
 539 YRNRFGNYIAQTLNDGRGPKSIEDDSEMKLVRYNQGADFYGAEGEYIFKPTPRYRIGV 598
 622 SGDVVRGRLKNLPSLPGREDAYGNRPPIAODDQNAQPRVPAARLGFHLKASLTDRIDANLD 681
 599 SGDVVRGRLKNLPSLPGREDAYGNRPPIAODDQNAQPRVPAARLGFHLKASLTDRIDANLD 658
 682 YRVRFAQNKARVETRTPGHHMLNLGANYRNRTRYGEWNNYVADNLLNQSVAHSSFLS 741
 659 YRVRFAQNKARVETRTPGHHMLNLGANYRNRTRYGEWNNYVADNLLNQSVAHSSFLS 718
 742 DTPQGRSFTGGVNVKF 758
 719 DTPQGRSFTGGVNVKF 735

RESULT 9
 ABP80923 standard; Protein; 728 AA.
 XX ABP80923;
 AC
 XX
 DT 07-MAR-2003 (first entry)
 XX
 DE N. gonorrhoeae amino acid sequence SEQ ID 8376.
 XX
 KW Antibacterial; infection; vaccine; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 XX
 FN WO200279243-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 12-FEB-2002; 2002WO-IB02069.
 XX
 PR 12-FEB-2001; 2001GB-0003424.
 XX
 FA (CHIR-) CHIRON SPA.
 XX
 PI Fontana MR, Pizza M, Massignani V, Monaci E;
 XX
 DR WPI; 2003-058415/05.
 DR N-PSDB; ABZ41893.
 XX
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 XX
 PS Disclosure; Page 799; 815pp; English.
 XX
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 XX
 SQ Sequence 728 AA;

Query Match 90.3%; Score 3644; DB 24; Length 728;
 Best Local Similarity 93.8%; Pred. No. 0;
 Matches 686; Conservative 11; Mismatches 28; Indels 6; Gaps 3;

31 VLETVTVVVGSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDGVPGIHASQYG 90
 1 VLETVTVVVGSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDGVPGIHASQYG 60
 91 GASAPVIRGQTRRIKVLNHHGEGTGMADFPDHAIMVDTALSQOVELRGPVTLTYSSG 150

61 GASAPVIRGQTRRIKVLNHHGEGTGMADFPDHAIMVDTALSQOVELRGPVTLTYSSG 120
 151 NVAGLVADVADGKIPEKMPENGVSQELGLRLSSGNLEKLTSGGINIGLQKPFVLTGELYR 210
 121 NVAGLVADVADGKIPEKMPENGVSQELGLRLSSGNLEKLTSGGINIGLQKPFVLTGELYR 180
 211 KSGDYAVPRYRNLRKLPDSRRFANGQHRV-LGW--RKFPYRRTYSDDRDQYGLPAHSH 267
 181 KSGDYAVPRYRNLRKLPDS---HADSTGSGISGLSVGEGKFGIAGYSDRRDQYGLPAHSH 237
 268 EYDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGHHDDDDAHAAHAKPKW 327
 238 EYDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGHHDDGGAHAHTHNGKPW 297
 328 IDLRNKRVELAEWKQPEPPFEALRVHLNRNDYHDEKAGDAVENFNNOTONARIELRH 387
 298 IDLRNKRVELAEWKQPEPPFEALRVHLNRNDYHDEKAGDAVENFNNKTHNARIELRH 357
 388 QPIGRLKSGWGVQVILGQKSSALSATSSEAVKQPMLLDNKVHVSFFGVEQANWNTLEGG 447
 358 QPIGRLKSGWGVQVILGQKSSALSATSSEAVKQPMLLDNKVHVSFFGVEQANWNTLEGG 417
 448 VRVEKOKASIRYDKALIDRENYVQPLDLGAHQRTARSFALSGNWTFTPOHKLSTLASH 507
 418 VRVEKOKASIRYDKALIDRENYVQPLDLGAHQRTARSFALSGNWTFTPHHKLSTLASH 477
 508 QERLPSTQELYAHGKHVATNTFEVGNKHLNKNERNIELALGYEGDRWQYNLALYRNRFG 567
 478 QERLPSTQELYAHGKHVATNTFEVGNKHLNKNERNIELALGYEGDRWQYNLALYRNRFG 537
 568 NYIYAQTLDNGRGPKSIEDDSEMKLVRYNQGADFYGAEGEYIFKPTPRYRIGVSGDYVR 627
 538 NYIYAQTLDNGRGPKSIEDDSEMKLVRYNQGADFYGAEGEYIFKPTPRYRIGVSGDYVR 597
 628 GRLKXLPSPGREDAYGNRPPIAODDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVRFA 687
 598 GRLKXLPSPGREDAYGNRPPIAODDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVRFA 657
 688 QNKLYRTRTPGHHMLNLGANYRNRTRYGEWNNYVADNLLNQSVAHSSFLSDTPQMG 747
 658 QNKLYRTRTPGHHMLNLGANYRNRTRYGEWNNYVADNLLNQSVAHSSFLSDTPQMG 717
 748 RSFTGGVNVKF 758
 718 RSFTGGVNVKF 728

RESULT 10
 AAY74562
 ID AAY74562 standard; Protein; 339 AA.
 XX
 AC AAY74562;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:598.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria meningitidis.
 XX
 FN WO957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.

```

PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR N-PSDB; AAZ53324.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX Claim 2; Page 419- ; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ54253 to AAZ55941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 339 AA;
XX
XX Query Match 44.8%; Score 1808; DB 21; Length 339;
XX Best Local Similarity 99.1%; Pred. No. 8.4e-157;
XX Matches 336; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 420 MLLDNKVQHYSPFGVEQANWNTLEGGVRVEKOKASIRYDKALIDRENYKQPLDLGA 479
XX Db 1 MLLDNKVQHYSPFGVEQANWNTLEGGVRVEKOKASIQYDKALIDRENYNHPPLDLGA 60
XX
XX QY 480 HROTARSPALSGNWFYTPQHKLSLTASHOERLPSTOELYAHGKHVATNTFEVGNKHLNKE 539
XX Db 61 HROTARSPALSGNWFYTPQHKLSLTASHOERLPSTOELYAHGKHVATNTFEVGNKHLNKE 120
XX
XX QY 540 RSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTINDGRGPKSIEDDSEMKLVRYNQSG 599
XX Db 121 RSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTINDGRGPKSIEDDSEMKLVRYNQSG 180
XX
XX QY 600 ADFYGAEGEYFKPTPRYRIGVSGDYVRGRLKNI.PSLPGREDAYGNRPFTAQDDONAPRV 659
XX Db 181 ADFYGAEGEYFKPTPRYRIGVSGDYVRGRLKNI.PSLPGREDAYGNRPFTAQDDONAPRV 240
XX
XX QY 660 PAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRNRTRYGEW 719
XX Db 241 PAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRNRTRYGEW 300
XX
XX QY 720 NWYKADNLLNQSVYAHSSFLSDTPQMGSRSTGGVNVKF 758
XX Db 301 NWYKADNLLNQSVYAHSSFLSDTPQMGSRSTGGVNVKF 339
XX
XX RESULT 11
XX AAY74563
XX ID AAY74563 standard; Protein; 339 AA.
XX AC AAY74563;
XX
XX DT 21-MAR-2000 (first entry)
XX

```

```

DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:600.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX antibacterial; gene therapy.
XX
XX OS Neisseria meningitidis.
XX
XX PN WO9957280-A2.
XX
XX PD 11-NOV-1999.
XX
XX PF 30-APR-1999; 99WO-US09346.
XX
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53325.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX PT vaccines and diagnostics -
XX
XX Claim 2; Page 420; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ54253 to AAZ55941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX SQ Sequence 339 AA;
XX
XX Query Match 44.4%; Score 1791; DB 21; Length 339;
XX Best Local Similarity 98.5%; Pred. No. 3.1e-155;
XX Matches 334; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX QY 420 MLLDNKVQHYSPFGVEQANWNTLEGGVRVEKOKASIRYDKALIDRENYKQPLDLGA 479
XX Db 1 MLLDNKVQHYSPFGVEQANWNTLEGGVRVEKOKASIRYDKALIDRENYNHPPLDLGA 60
XX
XX QY 480 HROTARSPALSGNWFYTPQHKLSLTASHOERLPSTOELYAHGKHVATNTFEVGNKHLNKE 539
XX Db 61 HROTARSPALSGNWFYTPQHKLSLTASHOERLPSTOELYAHGKHVATNTFEVGNKHLNKE 120
XX
XX QY 540 RSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTINDGRGPKSIEDDSEMKLVRYNQSG 599
XX Db 121 RSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTINDGRGPKSIEDDSEMKLVRYNQSG 180
XX
XX QY 600 ADFYGAEGEYFKPTPRYRIGVSGDYVRGRLKNI.PSLPGREDAYGNRPFTAQDDONAPRV 659
XX Db 181 ADFYGAEGEYFKPTPRYRIGVSGDYVRGRLKNI.PSLPGREDAYGNRPFTAQDDONAPRV 240
XX

```

QY 660 PAARLGPHLKASLDRIDANLDYRVFAQNKLYARVETPTGHHMLNLGANYRNTRYGEW 719
 DB 241 PAARLGPHLKASLDRIDANLDYRVFAQNKLYARVETPTGHHMLNLGANYRNTRYGEW 300
 QY 720 NWYVKADNLLNQSVAHSSFLSDTPQMGSRFTGGVNVKF 758
 DB 301 NWYVKADNLLNQSVAHSSFLSDTPQMGSRFTGGVNVKF 339

RESULT 12

ID AAY74561 standard; Protein; 339 AA.

AC AAY74561;

DT 21-MAR-2000 (first entry)

DE Neisseria gonorrhoeae ORF 149 protein sequence SEQ ID NO:596.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.

OS Neisseria gonorrhoeae.

XX WO9957280-A2.

PN 11-NOV-1999.

PD 30-APR-1999; 99WO-US093346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0098062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

DR N-PSDB; AA253323.

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

XX Claim 2; Page 418; 1453pp; English.

PS AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAY75941

CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 339 AA;

Query Match 43.7%; Score 1763; DB 21; Length 339;

Best Local Similarity 96.2%; Pred. No. 1.1e-152;
 Matches 326; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 420 MLIDNNVHYSHYFFGVEQANWDNFTLEGGRVVEKQKASIRYDKALIDRENYKQPLDGA 479

DB 1 MLIDNNVHYSHYFFGVEQANWDNFTLEGGRVVEKQKASIRYDKALIDRENYKQPLDGA 60

QY 480 HROTARSPALSGNWTFTPOHKLSTLASHOERLPSTQELYAHGKHVATNTFEVGNKHLNKE 539

DB 61 HROTARSPALSGNWTFTPHHKLSTLASHOERLPSTQELYAHGKHVATNTFEVGNKHLNKE 120

QY 540 RSNNIELALGYEGDRWQYNLALYRNFQNYIYAQTLNDGRGPKSIEDDSEMKLVRYNOSG 599

DB 121 RSNNIELALGYEGDRWQYNLALYRNFQNYIYAQTLNDGRGPKSIEDDSEMKLVRYNOSG 180

QY 600 ADPYGAEGEIVFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPPIAODDQNAPEV 659

DB 181 ADPYGAEGEIVFKPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPPIAODDQNAPEV 240

QY 660 PAARLGPHLKASLDRIDANLDYRVFAQNKLYARVETPTGHHMLNLGANYRNTRYGEW 719

DB 241 PAARLGPHLKASLDRIDANLDYRVFAQNKLYARVETPTGHHMLNLGANYRNTRYGEW 300

QY 720 NWYVKADNLLNQSVAHSSFLSDTPQMGSRFTGGVNVKF 758

DB 301 NWYVKADNLLNQSVAHSSFLSDTPQMGSRFTGGVNVKF 339

RESULT 13

AAO17586

ID AAO17586 standard; Protein; 814 AA.

XX AAO17586;

DT 19-JUL-2002 (first entry)

DE M catarrhalis MCA102062 protein SEQ ID NO: 52.

KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;
 KW auditory; antibacterial; otitis media; sinusitis; pneumonia.

XX Moraxella catarrhalis.

XX WO200218595-A2.

PD 07-MAR-2002.

PF 28-AUG-2001; 2001WO-CA01221.

XX 28-AUG-2000; 2000US-228294P.

PR 28-AUG-2000; 2000US-228295P.

PR 28-AUG-2000; 2000US-228296P.

PR 29-AUG-2000; 2000US-228438P.

PR 29-AUG-2000; 2000US-228439P.

PR 29-AUG-2000; 2000US-228440P.

PR 29-AUG-2000; 2000US-228441P.

PR 29-AUG-2000; 2000US-228442P.

PR 29-AUG-2000; 2000US-228443P.

PR 29-AUG-2000; 2000US-228512P.

PR 29-AUG-2000; 2000US-228512P.

PR 29-AUG-2000; 2000US-228742P.

PR 29-AUG-2000; 2000US-228773P.

PR 01-SEP-2000; 2000US-229465P.

PR 01-SEP-2000; 2000US-229474P.

PR 01-SEP-2000; 2000US-229475P.

PR 01-SEP-2000; 2000US-229478P.

PR 05-SEP-2000; 2000US-229740P.

PR 05-SEP-2000; 2000US-229803P.

PR 05-SEP-2000; 2000US-229804P.

PR 05-SEP-2000; 2000US-229805P.

PR 05-SEP-2000; 2000US-229806P.

PR 05-SEP-2000; 2000US-229809P.

PR 05-SEP-2000; 2000US-229811P.

PR 06-SEP-2000; 2000US-230214P.
 PR 06-SEP-2000; 2000US-230250P.
 PR 06-SEP-2000; 2000US-230252P.
 XX (AVET) AVENTIS PASTEUR LTD.
 XX Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;
 XX WPI; 2002-401721/43.
 XX N-PSDB; AAL46516.
 XX Moraxella polypeptide and polynucleotides useful as vaccine for
 PT immunizing a host e.g. humans against disease e.g. otitis media,
 PT pneumonia, caused by infection of the bacteria
 XX Claim 28; Fig 51; 277pp; English.
 XX The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarrhalis. These can be used to produce
 CC vaccines which protect against M. catarrhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a protein of the invention.
 XX Sequence 814 AA;
 SQ
 Query Match 39.2%; Score 1580.5; DB 23; Length 814;
 Best Local Similarity 41.2%; Pred. No. 2.6e-135;
 Matches 336; Conservative 141; Mismatches 270; Indels 69; Gaps 16;
 QY 7 KPIVLSIL-LINTPLLAQAH-ETQSVGLETVTVVVGSRPRATSGLLHTSTASDKIISGD 64
 DB 4 KPLACAILATFSMPMLAEANLKDPFTVLDGVSTSLADQNTFEGVNHSTVSGITVSKE 63
 QY 65 TLQKAVNLGALDGVPGIHAISQYGGGASAFVIRGQTRIKVNLNHGETGDMADSPDH 124
 DB 64 QLQQRATTLGALAGELGVHSHFGGASAFIRGQEGKELKILONGSEVVDMSGLSPDH 123
 QY 125 AIMVDTALSOQVEILRGPTVLLYSSGNVAGLVADGKIPEKMPENGVSGLGRLSSGN 184
 DB 124 AIAVDVTLAKOVELIRGSGALLYASGNSAGVNVVDKIPSKLPK-LQGDVTVRLSSAN 182
 QY 185 LEKLTSGGINILGKFNVLHTPEGLYRKSGDYAVPRY-----RNL 223
 DB 183 REKLITASAEAPLGHVAVRAGLSQAADYKTRFDRHVFNKXKEDDNTQPEFIYKDTL 242
 QY 224 KRLPDSPPRFANGQRAVLG--W--RKFPYRTYSDRRDQYGLPAHSHHYDDCHADLIWQ 279
 DB 243 KHLFDSHAK-----SNAGTGLVSWVGNQGLGASVSLRRDKYGLPNHSHYEYECVSHGISQ 298
 QY 280 KSLINKRYLOLYPHLLTBEVDYDNPGLSCGFHDD-----DDAHAAH-----NGKFWID 329
 DB 299 SALQYKFLRLYLPFLMENDDLLEFDNAGLECHTHDDHDHEDHDAHDEHDEHDKFWID 358
 QY 330 LRNKRYELRAEWKQPPFGFALRVLHNRNDYHHEKAGDAVENFNNQTNQARIELRHQP 389
 DB 359 LMKRYDVQGOINAPFAGIDIRASMGKVDYVHDEIDGGEKTSFFDNQANVRLEASHTP 418
 QY 390 I-----GRLKGSGVQYLQKSSAL-----SATSAVQKPMLLDNKVQHYFFGVEQAN-W 439
 DB 419 IHTPWKSGVFGVGYLTSKNSGLVPPRYEDGKNQDTONILHNKNTKGTGVFFFEYKPN 478
 QY 440 DNFTLEGGVRVEKQKASIRYDKALI-----DRENYKQPLED-----LGAHQRTAR 485
 DB 479 DKLTVDAAARKEQTITMDYDKDAIYQSLNLGLATAHEPDIRFKRLDSDGLNPKKQTAR 538
 QY 486 SFALSGNMYFTPOHKLSTASHOELPSTOBYLHAGHVAINTFEVGNKHLNKSNNIE 545
 DB 539 SYAVUTHLQLPKPKLSLNLHQBLPNAQBYLHAGHVAINTFEVGNKHLNKSNNID 598
 QY 546 LALYEGDRWYNIALYRNRRFNFIYAQTLND-GRGPKSIEDDSEMKLVRYNQSADFYG 604
 DB 599 LGLTFQGDKNYDLGGVHYDFDNVFLQTLQSYKQGLRGHEDKLTARYEQAAAFYG 658

QY 605 AEGEIVFKPTPRYRIGVSGDYVRGRLKXNLPSPGREDAYGNRPFIADQDNAPRVPARL 664
 DB 659 FDNVIGYQINDVYHVALFGDIVRGKLTNLDPKXGRTDAYGNRPFLIKQDSDHTPLPPKRL 718
 QY 665 GFHLKASLTDRIDANLDYVRVFAQNKLYARTTRPGHHMLNLGANYRNTY--GEWNWY 722
 DB 719 GMLKLTANTYANWSGFLRYRHTFKQDKLANFERPTPAHNLVNLGLNYQHKPSHQAGSVQVF 778
 QY 723 VKADNLLNQSYYAHSSFLSDTPQMGSRFTGGVNVKF 758
 DB 779 FNANLLNDKVFPAHETFFPDMPQMGSRNFMGLGANFKF 814
 RESULT 14
 AAB19945
 ID AAB19945 standard; Protein; 818 AA.
 XX AAB19945;
 AC AAB19945;
 XX 19-MAR-2001 (first entry)
 DT Moraxella catarrhalis BASB107.
 XX Moraxella catarrhalis BASB107.
 DE
 XX BASB107; infection; pneumonia; otitis media; sinusitis;
 KW therapy; diagnosis; vaccine; genetic immunization; antibacterial;
 KW antibiotic; receptor.
 XX OS Moraxella catarrhalis.
 XX PN WO2000071724-A2.
 XX PD 30-NOV-2000.
 XX 18-MAY-2000; 2000WO-EP04618.
 XX 24-MAY-1999; 99GB-0012038.
 PR 24-MAY-1999; 99GB-0012040.
 PR 01-JUN-1999; 99GB-0012674.
 PR 01-JUN-1999; 99GB-0012705.
 PR 02-JUN-1999; 99GB-0012838.
 PR 08-JUN-1999; 99GB-0013354.
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Thonnard J;
 XX WPI; 2001-025166/03.
 DR N-PSDB; AAA89216.
 XX New BASB103-108 polypeptides isolated from Moraxella catarrhalis
 PT bacterium, useful for diagnosing and producing vaccines against
 PT bacterial infections such as otitis media and pneumonia
 XX Claim 1; Page 74-75; 79pp; English.
 CC The present sequence is that of Moraxella catarrhalis ATCC 43617
 CC BASB107, a novel protein that shows amino acid sequence homology
 CC to Escherichia coli PhuE receptor precursor (outer membrane receptor
 CC for Fe(III)-coprogen, Fe(III)-ferritoxamine-B and Fe(III)-rhodotorulic
 CC acid). The invention provides M. catarrhalis BASB103-108
 CC polypeptides (see AAB19941-46) and polynucleotides (see AAA89212-17),
 CC and methods for producing the polypeptides by recombinant methods.
 CC Claimed vaccine compositions comprise a BASB103-108 polypeptide, or
 CC a BASB103-108 polynucleotide for genetic immunization. A claimed
 CC method for diagnosing a M. catarrhalis infection involves identifying
 CC a BASB103-108 polypeptide, or an antibody immunospecific to a
 CC BASB103-108 polypeptide, in a biological sample. Compositions
 CC comprising an immunologically effective amount of a BASB103-108
 CC polypeptide, or a polynucleotide encoding it, are used to generate
 CC an immune response in an animal. An antibody directed against a
 CC BASB103-108 polypeptide can be used to treat humans with M.
 CC catarrhalis disease. The polypeptides are also used as research
 CC reagents for the discovery of therapeutics and diagnostics useful

RESULT 15

[illegible]

XX Sequence 713 AA;
 Query Match 5.4%; Score 216; DB 22; Length 713;
 Best Local Similarity 21.6%; Pred. No. 2.7e-10;
 Matches 187; Conservative 103; Mismatches 298; Indels 278; Gaps 47;

QY 11 LSLILINTPL-LAQAHETEVSGLTETVTVGKSRPRATSGLLHTSTASDKLIISGDTLRQK 69
 DB 8 LSLLSLTAAAGFAHAENANVALDTVTVGDRQ-----GSKRTNIV---TLQOK 55

QY 70 ----AVNLGDALDGVPGIHASQYGGASAPVI-----RGOTGRRIKV-----LNHH 111
 DB 56 DESTATDMRELLKEEPSI---DFGGNGTSGFTLRGMGQNSVDIKVDNAYSQILYHQ 112

QY 112 GETGDMADFPDHAIMVDVTALSOQVEILLRQPVTLTYSSGNVAGLVADQKIPKMP-- 169
 DB 113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDAQ 153

QY 170 ---NGVSGELGLRLSSGNLEKLTSGGINIG---LGK--NFVLHTEGLY---RKSGDYAV 217
 DB 154 DLLKGLDKXWVRLNSG---FAGNNGSVYASVFGKEGNF---DGLFSYNRNDEKDYEA 206

QY 218 PR-YRNL---KRLPDSP-----RRFANGQHRVILGWRKPFYR--RTYSDRRDQY 260
 DB 207 GKGFNRVNGGKTVPYSDALKRSYLAKIGTTFGGDHRIVLSHMKDQHRGIETV---REEF 263

QY 261 GLPAHSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAH 320
 DB 264 AVGEN-----SRITIKRQAPAYRE--TTQSTNLTAYTKDGLFVEKLDANAY 309

QY 321 AHNGKPMIDLNRKRYELRAE---WKQPPGFPALRVHLNRNDYHDEKAGDAVENFNNQ 377
 DB 310 V-----LEKKYSADDKNGYAGNVKGNPNHTRATGCMNFNDSRLAE-----Q 353

QY 378 T-QNARIELRHQ---PIGRUKSGWGYVLGQKSSAL-----SATSEAVKQPMLLDNKVQ 427
 DB 354 TLLKVCINVRHDEIKPQAFINQPKIE---DKDATEEDKKGNREKIAKAYRLTNTPK 410

QY 428 HYSFPGVEQAN-WDNFTLEGGVVRVEKQASIRYDKALIDRENYYKQPLDILGAHRTARS 486
 DB 411 TDTGAVIEAIHEIDGTFTGLGLRVDYRFRVKTHTDGK-----TVSS 449

QY 487 FALSGNW--YFTPOHKLSTASHQ--ERLPSTQE-LYAHGKH---VATNTFEVGNKHLN 537
 DB 450 SSINFSFGVWQPREHWSFSASHNVASRFLYDALQTHKRGIIISADGT-----K 501

QY 538 KERSNNIELALGEGDRWQYNLALYR-----NRFNGYIYAQTLNDG-----578
 DB 502 ABRARNTGIFNVDGTFAANGSYFRQTIKDALANPQNRHDSVAVREAVNAGYIKNHGYE 561

QY 579 -----RGPKSLEDSEMKLVRYNSGADFYGAE-GEIY-----EKP 613
 DB 562 LGASYTGTGLTAKVGSRSKPRFYTHPKKLSANPE-----FGAQTGRTWTASLAYRFP- 616

QY 614 TPRYRIGVSGDYVRGLKNLPLPGREDAYGNRPFIADQDNAPRVPAAARGLFHILKASLT 673
 DB 617 NPNLEIGWRGYV-----QKATGSILAAGQKDRDGLNVVRVQGFVN----659

QY 674 DRIDANLDYRYPQAOKLARYETRFPGHMLNLGANYRNTRYGWNWYKADNLLNSV 733
 DB 660 -----DVFANWKPJGKOTLN-----VNLSVN-----NVDFREY 687

QY 734 YAHSSFLSDT-POMGRSFTGGVNVKF 758
 DB 688 YPHSQRTWNTLPGVGRDRLGVNYKF 713

RESULT 20
 ID ABP77241
 XX ABP77241 standard; Protein; 713 AA.
 AC ABP77241;

XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 1012.
 XX KW Antibacterial; infection; vaccine; gene therapy.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.
 XX PF 12-FEB-2002; 2002WO-IB02069.
 XX PR 12-FEB-2001; 2001GB-0003424.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
 XX DR WPI; 2003-058415/05.
 XX DR N-PSDB; ABZ38211.
 XX FT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX PT medicament for treating or preventing N. gonorrhoeae infection -
 XX PS Disclosure; Page 262; 815pp; English.
 XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
 XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
 XX CC antibodies that specifically bind to the proteins. The composition
 XX CC comprising the protein, nucleic acid or antibody is useful for the
 XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 XX CC infection, this may be in the form of a vaccine or gene therapy.
 XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 XX CC molecules of the invention.

QY Query Match 5.4%; Score 216; DB 24; Length 713;
 DB Best Local Similarity 21.6%; Pred. No. 2.7e-10;
 DB Matches 187; Conservative 103; Mismatches 298; Indels 278; Gaps 47;

QY 11 LSLILINTPL-LAQAHETEVSGLTETVTVGKSRPRATSGLLHTSTASDKLIISGDTLRQK 69
 DB 8 LSLLSLTAAAGFAHAENANVALDTVTVGDRQ-----GSKRTNIV---TLQOK 55

QY 70 ----AVNLGDALDGVPGIHASQYGGASAPVI-----RGOTGRRIKV-----LNHH 111
 DB 56 DESTATDMRELLKEEPSI---DFGGNGTSGFTLRGMGQNSVDIKVDNAYSQILYHQ 112

QY 112 GETGDMADFPDHAIMVDVTALSOQVEILLRQPVTLTYSSGNVAGLVADQKIPKMP-- 169
 DB 113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDAQ 153

QY 170 ---NGVSGELGLRLSSGNLEKLTSGGINIG---LGK--NFVLHTEGLY---RKSGDYAV 217
 DB 154 DLLKGLDKXWVRLNSG---FAGNNGSVYASVFGKEGNF---DGLFSYNRNDEKDYEA 206

QY 218 PR-YRNL---KRLPDSP-----RRFANGQHRVILGWRKPFYR--RTYSDRRDQY 260
 DB 207 GKGFNRVNGGKTVPYSDALKRSYLAKIGTTFGGDHRIVLSHMKDQHRGIETV---REEF 263

QY 261 GLPAHSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAH 320
 DB 264 AVGEN-----SRITIKRQAPAYRE--TTQSTNLTAYTKDGLFVEKLDANAY 309

QY 321 AHNGKPMIDLNRKRYELRAE---WKQPPGFPALRVHLNRNDYHDEKAGDAVENFNNQ 377
 DB 310 V-----LEKKYSADDKNGYAGNVKGNPNHTRATGCMNFNDSRLAE-----Q 353

QY 378 T-QNARIELRHQ---PIGRUKSGWGYVLGQKSSAL-----SATSEAVKQPMLLDNKVQ 427

Db 354 TLLKYGINRHOEIKPQFLNSQFKIE---DKDATEDKKRRENEKIKAAYRLTNPTK 410
 Qy 428 HYSFEGVEQAN-WDNFTLEGGVVRVEKQKASIRYDKALIDRENYYKQPLDGAHRQTARS 486
 Db 411 TDTGAYIEAIIHEIDGFTLTGLRYDRFKVTHDGK-----TVSS 449
 Qy 487 FALSGNW--YETPOHKLISLTASHQ--ERLPSTQE-LYAHGKH-----VATNTEVGNKHLN 537
 Db 450 SSLNFSFGVIVQPREHWSFSASHNVASRPDYDALQTHGKKGIIISADGT-----K 501
 Qy 538 KERSNNIELALGYEDRWQYNLALYR-----NRFNGYIYAQTILNDG----- 578
 Db 502 AEARANTEIGFNYNDGTFAANGSYRQTIKDALANPQNEHDSVAVREAVNAGYIKNHGVE 561
 Qy 579 -----RGPKEIDSEMKLVRYNQSADFYGAE-GEIY-----FKP 613
 Db 562 LGASYRTGGLTAKVGVSRSKRFYDTHPKLLSANPE-----FGAQTGRWTASLAYREF- 616
 Qy 614 TPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFIADDDQNAVRPAARLGFHLKASLT 673
 Db 617 NPNLEIGWRGRVY-----QKATGSILAAQKQKRDGKLENVVRQFGVY----- 659
 Qy 674 DRIDANLDYRVRFAQNKLYARETRTPGHHMLNMGANYRNRTRYGEWNVYVRADNLLNQSV 733
 Db 660 -----DVFAWKPLGKDTLN-----VNLSVN-----NVFDREY 687
 Qy 734 YAHSSFLSDT-PQMGRSFTGGVNVKF 758
 Db 688 YPHSQRWNTLTGVRGRDVRGLGVNVKF 713

RESULT 21

ID ABB78069 standard; Protein; 720 AA.
 AC ABB78069;

DT 05-NOV-2002 (first entry)
 DE Amino acid sequence of p64 polypeptide.

KW p177; p88; p64; p55; p46; vaccine; gonorrhea.
 OS Neisseria gonorrhea.

PN W0200260936-A2.

PD 08-AUG-2002.

PF 31-JAN-2002; 2002WO-US02881.

PR 31-JAN-2001; 2001US-266070P.

PR 06-AUG-2001; 2001US-310356P.

PR 23-OCT-2001; 2001US-344452P.

XX (IOWA) UNIV IOWA RES FOUND.

PA (REGC) UNIV CALIFORNIA.

PA (APIC/) APICELLA M. A.

PA (EDWA/) EDWARDS J L.

PA (GIBS/) GIBSON B W.

PA (SCHE/) SCHEFFLER K.

PA (BROW/) BROWN E.

PI Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;

DR WPI; 2002-619227/66.

DR N-PSDB; ABQ78300.

XX

PT New polypeptide comprising p177, p88, p64, p55 or p46 from Neisseria
 PT gonorrhea, useful for preventing, or protecting a female patient
 PT against, N. gonorrhea colonization or infection -

XX

PS Claim 7; Page 117-120; 130pp; English.

CC The present sequence represents a p64 polypeptide. The specification
 CC describes p177, p88, p64, p55 and p46 polypeptides from Neisseria
 CC gonorrhea. The polypeptides are useful as vaccines, for preventing,
 CC or protecting a female patient against, N. gonorrhea colonization or
 CC infection. Such immunisation can prevent gonorrhea in women.

SQ Sequence 720 AA;

Query Match 5.3%; Score 214.5; DB 23; Length 720;
 Best Local Similarity 21.6%; Pred. No. 3.8e-10;
 Matches 186; Conservative 101; Mismatches 315; Indels 259; Gaps 45;

Qy 16 INTPL-----LAQAHETEQSVGLTETVVGKSRPRATSGLIHTTASDKITS 62
 Db 1 MNTPLRLSLLSLTLAAGFAHAENNAKVVLDTVTVKGDRQ-----GSKIRTNIV- 50
 Qy 63 GDTLRQK-----AVNLGDALDGVPGIHASQYGGASAPVI-----RGQTGRRIKV----- 107
 Db 51 --TLOOKDESTATDMRELLKBEPSI---DFGGNGTSQFLTLRGMGQNSVDIKVDNAYSD 105
 Qy 108 ---LNHHGETGDMADFPDHALIWDLTALSOQVEILRGPTVLLYSNGVAGLVADVADGKPT 164
 Db 106 SQILLHQGR-----FIVDPALVKVSVQVK-----AGSASAGIGATNGAII 146
 Qy 165 EKMPK-----NGVSGELGLRLSSG--NLEKLTSGGINIGLGNFVLHTGELY---RKSG 213
 Db 147 AKTVDAQDLLKGLDKNWGVRLNSGFASNEGSYGASVFGKEGNF---DGLSYNRNDEK 202
 Qy 214 DYAVPR-YRNL---KRLPDSP-----RRFANGQRAVLGWRKFRFYR----- 251
 Db 203 DYEAGKGFNVNGKTVFPYSALDKRSYLAKIGTTFGDDDHRIVLSHMKDQHRGIRTVREE 262
 Qy 252 -TYSRRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEDEVDYDNGLSCG 310
 Db 263 FTVGKSSRINIDROQAPAYR-----TTQSTNLTAYTKNLG 299
 Qy 311 FHDDDDAHAAHANGKPIDLRNKRY---ELRAEMKQPPFGPEALRVHLNRNDYHDEKAG 367
 Db 300 FVEKLDANAYV-----LEKERYSADDSGTGVAGNVKGNHTRITTRGANFNFDRLA 351
 Qy 368 DAVENFFNQOT-QNARIELRHQ---PIGRLLKSGWGVYLGQKSSALSATSEAVKPM--- 420
 Db 352 E-----QTLKYGINYRHOEIKPQAFNLNKSFTPTTEKN-----GQKVDKPMEQ 397
 Qy 421 -----LLDNKVQHVYFFGVEQA-NWDNFTLEGVVRVEKQKASIRYDKALIDR 466
 Db 398 MKDRADEDTHAYKLSNPTKTDGVVVEAIIHIDIGDFTLTGLRYDRFKVK-THDGKTVSS 456
 Qy 467 ENYKQPLPLDGAHRQTARSFALSGNWYFTPQHKLSLTASHQERLPSTQELVAHGKH--- 523
 Db 457 SNLN-----PSFGVIWQPE-----HWSFSASHNY---ASRSPRL--YDALQTHGKRII 501
 Qy 524 -VATNTEVGNKHLNKNRSNNIELALGYEDRWQYNLALYRNRFGNYIYAQTLDNGRGP 582
 Db 502 SIADGT-----KABEARNTEIGFNYNDGTFAAN-----GSYEW-QTIKDALANP 544
 Qy 583 SIEDDSEMKLVRYNQSADFYGAE-GEIYFKPTPRYRIGVSGDYVRGRLKNLPSLPGRED 641
 Db 545 QNRHDSVAVREAVNAGYIKNHGYELGASFTYGGTAKVGVS-----HSKPRFYD 593
 Qy 642 AYGNRPFIAQDDQNAVRPAARLGFHLKASLTDRI-DANLDY-VRVFAQNKLYARETRTP 699
 Db 594 THDKLLSANPER-----GAQVGRWTASLAYRFPQNPLEIGHR-----GRYVQKAT 640
 Qy 700 GHEMLNLGANYRR-----NTRYGEW-----NMVYKADNLLNQSVYAHSS 738
 Db 641 G-SILAAQKQKRDKNLENVVRKGFVNDVFANWPKLGDTLNVLNVNFKYFYHPSQ 699
 Qy 739 FLSDT-PQMGRSFTGGVNVKF 758
 Db 700 RWTNLTGVRGRDVRGLGVNVKF 720

RESULT 22
AAW01462
ID AAW01462 standard; Protein; 719 AA.
XX
AC AAW01462;
XX
DT 24-FEB-1997 (first entry)
XX
DE NTHI HxuC protein.
XX
KW HxuC; NTHI; vaccine; genetic immunisation; diagnosis;
XX meningitis; pneumonia; bacteraemia; otitis media.
XX
OS Haemophilus influenzae nontypeable strain N182.
XX
XX Key Location/Qualifiers
FH 6..18
Region /label= TonB box
FT Cleavage-site 19..21
FT /note= "signal peptidase I cleavage site"
XX
XX WO9633275-A1.
XX
XX 24-OCT-1996.
XX
XX 15-APR-1996; 96WO-US05167.
XX
XX 20-APR-1995; 95US-0425843.
XX
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX Cope LD, Hansen EJ, Hanson MS, Jarosik GP;
XX
XX WPI; 1996-485781/48.
XX
XX N-FSDB; AAT44520.
XX
XX Genes encoding H. influenzae HxuC and HxuD surface-expressed
PT protein(s) - useful in the prep. of vaccines for children against
PT H. influenzae infection
XX
XX Claim 11; Page 153-159; 188pp; English.
XX
XX The HxuC protein (AAW01462, see also AAW01461) of nontypeable
CC Haemophilus influenzae (NTHI) strain N182 is a 78 kDa outer
CC membrane protein involved in the utilisation of low levels of free
CC haem. It shows homology to TonB-dependent outer membrane proteins
CC of other bacteria, and was identified from an open reading frame
CC (see also AAT44520) isolated by PCR amplification. Another outer
CC membrane protein, HxuD (AAW01465, see also AAW01464), was similarly
CC identified. Recombinant HxuC and HxuD proteins can be prep. in
CC transformed host cells and used: to prepare vaccines against NTHI
CC infection; to raise diagnostic antibodies; and to prepare NTHI
CC diagnostic or therapeutic compans.
XX
XX Sequence 719 AA;
SQ
Query Match 5.3%; Score 214; DB 17; Length 719;
Best Local Similarity 19.6%; Pred. No. 4.2e-10;
Matches 158; Conservative 133; Mismatches 286; Indels 230; Gaps 42;
XX 11 LSLALINTPLLAQAHTETEQVGLTVVVGKSRPRATSGLLHTSTASDKIISGDTLROKA 70
DB 6 LSLAIITTT-LVTANALAQSVELDSINVITATRDP---SRFAYTPEKQSK---DSLSSKQA 57
XX 71 VNLGDMALDGVPGIHASQYGGG---ASAPVIRQGTGR-IKVLNHGFTGDMADSPDH-- 124
DB 58 TSVAALAEIPNVDR--CGSRSIAQKENIRGLSDNRVQVVDIGVRQNFDLA-----HRG 110
XX 125 AINVDPTALSQVEILRGPTVLLYSSGNVAGLVVDADGKIPEKMPENGVSGELGLRLSSG- 183
DB 111 SYFLPMSLIQIEVIEKGPSSSLWGSGLGCVVAMTFNALDLLKNN---DRFGVKIRQGY 167

QY 184 -NLEKLTSGGINI-GLGKNFVLHTEGLYRKSGDYAVPRYRNKLKRLPDSRRFA----- 234
DB 168 QTANNLSERDVSVAANDKFDVLISGFYNADNLRTGNGKLN--TAYKQFGGLAKFGW 225
QY 235 --NGQHRAVLGWRKEFYRRTYSRRDQVGLPAHSEYDDCHADIIWOKSLINKRYLQIYP 292
DB 226 SINDANRVLSHRETRFKQT-----APSNNEV-----ENELTNKQ----- 260
QY 293 HLLTEEDVDYDNPGLSCGFHDDDDAHAAHNGKWPIDLNRKRYELRAEWKPPFGFEALR 352
DB 261 --ITDQ1BEYH-----RSKXTKPSLE-----RVEKQKA-- 286
QY 353 VHLNRNDYHDEKAGD--AVENFF---NNOTQNAIELRHQPIGLRLKSGWGVYLGQKSS 407
DB 287 TRFGSVSYLSQQIPDQSTVFNYVLTDPNVLNTHIALYNNKT-----IEKEQRKVS 338
QY 408 ALSATSEAVKQPMILLDN--KVQHYSP-FGVEQANWNTLEGGV-----RVEKQKA-- 455
DB 339 GVKDQTKLTTRGILNRNSSELSHISFVYGVDMR-DKIRTERGTNSSDAKFRAPYNANS 397
QY 456 -----SIRYD-----KALIDRENYIKQPLDPLGAHRQTARSF 487
DB 398 NTTGVYLLAHIPLFGEKLLSPSVYDHYDTSSKTVKYKDNHL-----SP 442
QY 488 ALSGNWYFTPOHKLISLTASHOE--RLPSTQELYAHGKV-----ATNTFEVGNKHL 536
DB 443 ATKLTWIVT--NWLDFTAKYNEAFRPSQMERFVSGAHFGASIIIGLNAINRF-VANPNL 499
QY 537 NKERSNNIEL-----ALGYEGDRWQYNLALYRNFGNYIAQTLNDGRGPKSIEDDSE 589
DB 500 RPETAKNKEITANLHFDLSLFKQGRKIEATYFRDNDVKDFINLKIFDANTNK----- 552
QY 590 MKLVRYNQSADFYGABGEIYFKPT-PRYR-----IGVSGDYVRGRL---KNLPSL 636
DB 553 -----NASASASASANGTLLPLPTKQYONITNARLSGIELQAQYQTERLTLFTNYGST 606
QY 637 PGREDAYGNRPFLAQDDQNAVRPAARLGFHLKASLTD---RIDANLDYRYRFAONKLAR 693
DB 607 KGKDKDSG-----EALSNAASKIGVGVNYALVKDKFTVGATVTTHYA--AORRVPK 655
QY 694 YETRT-PGHHMLNLGANYRRNTRYGEW 719
DB 656 DHSVTYPSVILTDLRATYA--PLKGEW 680
RESULT 23
AAW75567
ID AAW75567 standard; Protein; 703 AA.
XX
AC AAW75567;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 761 protein sequence SEQ ID NO:2608.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
XX WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 XX WPI; 2000-062150/05.
 DR N-PSDB; AAZ54329.
 XX
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 PT
 XX
 XX Claim 2; Page 1237; 1453pp; English.
 XX
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 XX Sequence 703 AA;
 SQ

Query Match 5.3%; Score 212.5; DB 21; Length 703;
 Best Local Similarity 20.7%; Pred. No. 5.6e-10;
 Matches 182; Conservative 117; Mismatches 260; Indels 321; Gaps 52;

QY 12 SILLINTPLA---QAHETEOSVLETVTVGKSRPRATSLH---TSTADKIISG- 63
 DB 12 TLIIASFVAADTQDNGEHTALTPTVSVVGGQSDTVLKGINIDYDEAAVTRNGQLIKET 71
 QY 64 ---DTLR-QKAVNLG---DALDGVPGIHASQVGGASAPVIRGOTGRRRIKVLNHGE 113
 DB 72 PQTIDTLNIQKNVGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGP 117
 QY 114 TGDWADSPDHA-----IMVDTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPKMP 168
 DB 118 QADASDIYRDGVRESGQVRRSTANIERVEILKGFSSVLYGRTNGGGINVMVSKYANFKQS 177
 QY 169 EN--GVSG-----ELGLRLSS--GNLEKLTSG--GINIGLGNFVLIHT 205
 DB 178 ENIGAVYGSWANRSLNMDINEVLNKNVAIRLTGEVGRANSFRSGIDSKNVWVPSIIVKL 237
 QY 206 EGLYKSGDYAVPRYLNKRLPD--SPRFANGQRAVLGWRKFRYRTYSDRRDQYGLP- 263
 DB 238 DNLGKWTGQYT---YDNVETPDRSP-----TKSVY-----DRFGLPY 272
 QY 264 ---AHSHEYDDCHADIIWQSLINKRYLQLYPHLLTBEEDVDYDNPGLSCGFHDDDDAHA 319
 DB 273 RMGFAHRNDF-----VKDKLQW-----RSLEY----- 296
 QY 320 HAHNGKPTIDLRNRYELRAEWKQPPGFALRVHLNRNDYHH---DEKAGDAVENFF-- 374
 DB 297 -AFNDK-W-----RAQWQ-----LAHRTAQDPDFHFAVSGENLIKRNVAW 336
 QY 375 ---NNQTONARIELR-HQPIGLKSG--SWGQVYLGQK-----SSALSAT----- 412
 DB 337 QQTDNKTLSSNLTNGDYTIGRFENHLLTVGMDYSGREHNPTLGFSSAFSASINPYDRASW 396
 QY 413 -SEAVKQWMLDN--KVQHYSPF--GVFQANWD--NFTLEGGRVREKQKASIRYDKALIDR 466

DB 397 PASGRQLPILQNRHKADSYGIFVONIFSATPDLKFVJGG-----RYDKYTFNS 445
 QY 467 ENYRQPLPDLGAHRQ--TARSFA--LSCNMYFTPOHKLSTASHOERLPSTQELYAHOKH 523
 DB 446 ENKL-----TGSSRQYSGHSFSPNIGAVMINPVH--TLVAS"NK--GPAPYGGRGY 494
 QY 524 VATNTFEVGNKHLNKRNSNNIELAL--GYEGDRWQYNLALYRNRFNPGNVIYAQTLDNDRGP 581
 DB 495 LSIDTLSSAVFNADPEYTRQYETGVKSSWLDRLSTLSAY----- 535
 QY 582 KSIEDDSEMLKVRYNQSGADFYGABGEIYFKPTPR---YRIGVSGD----- 624
 DB 536 -----QIERFN-----IYRPDPKNPNYIYAVSGKHSRGSVELSAIGOI 574
 QY 625 -----YVRGRKMLPSLPCREDAYGNRRPFIADDDQNAVRPAARLGFHLKASLTDRIDAN 679
 DB 575 IPKLYLRGLSGVMQAK-----VVEDKENPDV-----GIHL--NNTSNVTGN 615
 QY 680 LDYRV-----FAQNKLAHYETRTPGHMLN--LGANYRR-NTRYGE 718
 DB 616 L--FFRYTPPENLYGEIGVTGKRYGASRNKEVTTLPGFARVDAMLGNHKNVNVTF-- 673
 QY 719 MNWYVKADNLLNQSVYAHSSFLSDTFQMGSRSTGGVNVKF 758
 DB 674 -----AANLANQKYWRSDSM---PGNPRGYTARVNYRF 703

RESULT 24
 AAB68927
 ID AAB68927 standard; Protein; 703 AA.
 XX
 XX AAB68927;
 XX
 DT 18-APR-2001 (first entry)
 XX
 XX Neisseria meningitidis protein #26.
 XX
 KW Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
 KW rni5, rth; tolC.
 XX
 OS Neisseria meningitidis.
 XX
 PN EF1069133-A1.
 XX
 PD 17-JAN-2001.
 XX
 PF 13-JUL-1999; 99EP-0401764.
 XX
 PR 13-JUL-1999; 99EP-0401764.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Nassif X, Tinsley C;
 XX
 DR WPI; 2001-082916/10.
 DR N-PSDB; AAF56467.
 XX
 PT Immunogenic polypeptides derived from Neisseria meningitidis and the
 PT nucleic acids that encode them, useful for diagnosing and vaccinating
 PT against Neisseria infections e.g. bacteraemia and meningitis -
 XX
 PS Claim 3; Fig 26B; 240pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC several genes from Neisseria meningitidis. These include the dsbA, fhaB,
 CC fhuA, rni5, rth17, rth18, rth19, rth20, rth21 and tolC genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium,
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.
 XX
 SQ Sequence 703 AA;

36 WILLIAM

QY	352	RVHLNRNDYHDEKAGDAVENFFNNQTONARIELRHQPIGRKLGSGWGYLQOKSSALSA	411
Db	358	KYGIN---YRHOETKPAQ---FLNSQ---FKIEKKEKATDEEKNK-----NR	395
QY	412	TSEAVKQPMLLDNKVQHVSPFGVEQAN-WDNFTLEGGVRVEKOKASIRYDKALIDRENY	470
Db	396	ENEKIATAYRLTNPTKTDGTAYIEAIEHIDGFTLTGLGYDRFKVK-THDGKTVSSNNLN	454
QY	471	KQPLDPLGAHRTARSPALSNWYFTFPQHKLSTASHQERLPSTOELYAHGXH----	526
Db	455	-----PSFGVIMQPHB-----HWSFSASHNY---ASRSPRL--YDALQTHGKRGII	499
QY	527	NTFEVGNKHLNKRNNNTEIALGVEGDRWQVNUALYNNRFGNYIYAQTLDG-RGPKSIE	585
Db	500	GT-----KAERARNTTEIGFNYNDGTFFAAN-----GSYFW-QTIKDALANPQNRH	542
QY	586	DDSEMK-----LVRYNQSADYFGAEGEYTFKPTPRVRI	619
Db	543	DSVAVREAVNAGYIKQHGVELGASYRTGGLTAKVGVSHSKPRFYDTHKOKLLSANPEGA	602
QY	620	GVSGDYVRG---RLKNLPSLP-----GR--EDAYGNRPFIAQDDONAPRVPAPRLGHLKA	670
Db	603	QVGRTWTASLAYRFQN-ENLBIWGRGYVQKAVGSILVAGQKDRNGKLENVVRKGFQVNV-	660
QY	671	SLTDRIDANLDYRVFAQNKILARYETPTPGHHMLNLGANYRNTRYGEMNYYVKADNLLN	730
Db	661	-----DVFAWKPGLKDTLN-----VNLSVN-----NVFN	685
QY	731	QSVYAHSSFLSDT-PQMGSRFTGGVNVKP	758
Db	686	TFYPHQSQRWNTLFGVGRDVRLGVNYKPF	714
RESULT 26			
AAAY75568			
XX	ID	AAAY75568 standard; Protein; 703 AA.	
XX	AC	AAAY75568;	
XX	XX		
DT	21-MAR-2000	(first entry)	
XX	XX	Neisseria meningitidis ORF 761 protein sequence SEQ ID NO:2610.	
DE	DE	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;	
KW	KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;	
KW	KW	antibacterial; gene therapy.	
XX	XX	Neisseria meningitidis.	
OS	XX		
XX	XX	WO9957280-A2.	
FN	XX	11-NOV-1999.	
PD	XX		
PF	XX	30-APR-1999; 99WO-US09346.	
XX	XX		
XX	XX	01-MAY-1998; 98US-0083758.	
PR	31-JUL-1998;	98US-0094869.	
PR	02-SEP-1998;	98US-0098994.	
PR	02-SEP-1998;	98US-0099062.	
PR	09-OCT-1998;	98US-0103749.	
PR	09-OCT-1998;	98US-0103794.	
PR	09-OCT-1998;	98US-0103796.	
PR	25-FEB-1999;	99US-0121528.	
XX	XX		
PA	(CHIR)	CHIRON CORP.	
PA	(GENO-)	INST GENOMIC RES.	
XX	XX		
XX	XX	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;	
PI	PI	Petersen J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;	
PI	PI	Tettelin H, Venter JC;	
XX	XX	WPI; 2000-062150/05.	

DR	N-PSDB; AAZ54330.	Novel Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics -	Claim 2; Page 1238; 1453pp; English.
XX	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941	represent novel <i>Neisseria meningitidis</i> and <i>N. gonorrhoeae</i> polynucleotides and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent PCR primers used in the exemplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to <i>Neisserial</i> bacteria (e.g. <i>meningitis</i> and <i>septicaemia</i>), to detect the presence of <i>Neisseria</i> bacteria, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.	
XX	Sequence 703 AA;		
XX	Query Match 5.2%; Score 208.5; DB 21; Length 703;		
XX	Best Local Similarity 20.6%; Pred. No. 1.3e-09;		
XX	Matches 181; Conservative 119; Mismatches 259; Indels 321; Gaps 52;		
CC	12 SILLINTPLLA---QAHTEQSVGLETVTVCKSRPRATSGLLH---TSTASDKLIIG- 63		
CC	12 TLIIASFVPAADTQNGEHYATLPTVSVWGSDTSVLKGVINDEAAVTRNGQILKET 71		
CC	64 ---DTLR-QKAVNLG---DALDGVPCIHASQYGGGASAPVIRGOTGRRIKVLNHHGE 113		
CC	72 PTDITLAIQKNKYGTNDLSILEGNAGIDA-----AYDMRGES-----IFLRGF 117		
CC	114 TQDMADFPDHA-----IMVDTALSQQVEILRGPTVTLKSYSGNAGLVADVADKITEKMP 168		
CC	118 QADASDIYRDGVRESGQVRSSTANIERVELKGPSVLYGRTNGGGVIMVNSKYANFKQS 177		
CC	169 EN-----GVSGELGLRLSS--GNLEKLTSG--GINIGLGNFVLHT 205		
CC	178 RNIGTVGSMANRSLNMDINEVLNKNVALRLTGEVRANSFRSGIDSKNWNVSPSITVKL 237		
CC	206 EGLYRKSGDYAVPRYRNLRPLD--SPRRPANGQRAVLGWRKRFYRRTYSDRDQYGLP- 263		
CC	238 DNLGKWTGQYT--YDNVERTDPRSP-----TKSVY-----DRFLGPLY 272		
CC	264 ----AHSHEYDDCHADI1QWKLINRKYLYLPHLLTEEDVDYONPGLSCGFHDDDAHA 319		
CC	273 RMGFARHNDP-----VKDKLQW-----RSDLEY----- 296		
CC	320 HAHNGKPMWIDLRNKRYELRAEWKQPPGPFPEARLVHLNRNDYIH---DEKAGDAVENFF-- 374		
CC	297 -AFNDK-W-----RAQWQ-----LAHRTAAQDFDHFYAGSENGNLIKRYAW 336		
CC	375 ---NNQQTQWARIELR-HQDIGHLKG--SWGQVYLGQK-----SALSAT----- 412		
CC	337 QQTNDKTLSSNLTLAGDXYTIGFENHLYTGMDSYREHNPFTLGFSSAFSASINPYDRAW 396		
CC	413 -SEAVKQPMLLDN--KVQHSYSPF--GVEQANWD-NFTLEGGRVEREKASIRYDKALIDR 466		
CC	397 PASGRLOPILTORRHKADSYGIVQNIFSATPDLKEVLG-----RVDKTYFNS 445		
CC	467 ENYKQPLDPLGAHQ- -TARSPA--LSGNWYFTPHKLSLTASHQBRPLPSTOELYAHGKH 523		
CC	446 ENKL-----TGSSRQYSGHSPSPNIGAVWNNIPVH--TLVASYNK--GFAPYGGRGY 494		
CC	524 VATNTPFVGNKHLNRSNNIELAL--GYEGDRWQYNLALYRNPFGNYIYAQTINDGRGP 581		
CC	495 LSIDTLSSAVFNADPEYTRQYETGVKSSWMLDRLSTLTSAY-----YRIGVSGD- 535		
CC	582 KSIDDESEMKLVRYNQSGADFYCAEGEYIFKPTPR---YRIGVSGD----- 624		

Db	536	-----QIERFN-----	IRYDPKNNPIYAVSGKSRGVELSALGQI 574
Qy	625	-----YVRGLKNLPSLPGRDAYGNRPPIAQDQNAAPRVAARLGFHLKASLTDRIAN 679	
Db	575	IPKLYLRGSLGVWQAK-----	VVEDKENPDRV-----GIHL--NNTSNVTGN 615
Qy	680	LDYR-----	VFAQNKLYARET-----PGHMLN--LGANYR--NTRYGE 718
Db	616	L-FRYTPTENLYGEIGVTGKRYGYSRKEVITLPGFARVDAMLGNHKNVNTFA- 673	
Qy	719	WNWVKADNLNLSQVVAHSSFLSDTPQMGSRFTGGVNVKF 758	
Db	674	-----AANLFNQKYWRSDSM-----	PGNPGYTARVNYRF 703
RESULT 27			
ID	AA68916	standard; Protein; 703 AA.	
AC	AA68916;		
DT	18-APR-2001	(first entry)	
DE	Neisseria meningitidis protein #15.		
KW	Meningococcus; meningitis; bacteraemia; vaccine; dsba; fhaB; fhuA;		
KW	rn15; rth; tolC.		
OS	Neisseria meningitidis.		
PN	EP1069133-A1.		
PD	17-JAN-2001.		
PF	13-JUL-1999; 99EP-0401764.		
PR	13-JUL-1999; 99EP-0401764.		
PA	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.		
PA	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.		
PI	Nassif X, Tinsley C;		
XX	WPI; 2001-082916/10.		
DR	N-PSDB; AAF56456.		
PT	Immunogenic polypeptides derived from Neisseria meningitidis and the		
PT	nucleic acids that encode them, useful for diagnosing and vaccinating		
XX	against Neisseria infections e.g. bacteraemia and meningitis -		
PS	Claim 3; Fig 15B; 240pp; English.		
CC	The present invention provides the protein and coding sequences of		
CC	several genes from Neisseria meningitidis. These include the dsba, fhaB,		
CC	fhuA, rn15, rth17, rth18, rth19, rth20, rth21 and tolC genes. These can		
CC	be used in the diagnosis and treatment of infection by the bacterium,		
CC	which can lead to meningitis and bacteraemia, and in vaccines to prevent		
CC	such infection.		
XX	Sequence 703 AA;		
Qy	Query Match	5.2%; Score 208.5; DB 22; Length 703;	
Db	Best Local Similarity	20.6%; Pred. No. 1.3e-09;	
XX	Matches 181; Conservative 119; Mismatches 259; Indels 321; Gaps 52;		
Qy	12 SILLINTPLA---QAHETEQVGLTIVVVKSRPRATSGLLH-----TSTASDKIIG- 63		
Db	12 TLIIAFPPAAADTQNGEYHATATPTVSVGQSDTSVLKGYINDYDEAAVTENGQLIKET 71		
Qy	64 ---DTLR-QKAVNLG-----DALDGVPGIHASQVGGGASAPVIRGOTGRRIKVLNHHGE 113		
Db	72 PQTIDTLNQKKNKYGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGF 117		
Qy	114	TGDMADFPDHA-----IMVDTALSQQVEILRLPVTLLYSSGNVAGLVVDVADGKIPERP 168	
Db	118	QADASDIYRDGVRESGVRRSTRANIERVEILKGPSSVLYGRNGGVVIMVSKYANFQS 177	
Qy	169	EN-----GVSGELGRUSS--GNLEKLTSG--GINIGLGNFVLHT 205	
Db	178	RNIGTVYGSWANRSLNMDINEVLNKNVAIRLTGEVGRANSFRSGIDSKNVVSPSITVKL 237	
Qy	206	EGLYRKSGDYAPRYRNLKRLPD--SPRFANGQHRVLAIGWRKFRYRRTYSRRDOYGLP- 263	
Db	238	DNGLKWTGOYT---YDNVERTDRSP-----TKSVY-----DFGLPY 272	
Qy	264	-----AHSHEYDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSCGHHDDDAHA 319	
Db	273	RMGFAHRNDF-----VKDKLQVW-----RSDLEY----- 296	
Qy	320	HAHNGKPIDLNRKRYELRAEWKQPPGFPEALRVHLNRNDYHH---DEKAGDAVENFF-- 374	
Db	297	-AFNDK-W-----RAQWQ-----LAHTAAQDFDHFYAGSNGNLIKNTAW 336	
Qy	375	---NNOTONARIELR-HQPIGRKLG--SWGVOYLQOK-----SSALSAT----- 412	
Db	337	QQTDNKTLSSNLTNGDYITIGRFENHLTVGMDYSREHRNPTLGFSSAFSASINPYDRASW 396	
Qy	413	-SEAVKQPMLLDN--KVOHYSFF--GVEQANWD--NFTLEGVVRVEKQKASIRYKALIDR 466	
Db	397	PASGRLOPILTQNRHKADSYGIFVQNFISATPDLKFLVGG-----RYDKYTFNS 445	
Qy	467	ENVYKQPLPDGAHRO--TARSPA--LSGNWYFTPOHKLSTASHOERLPSTQELYAHGKH 523	
Db	446	ENKL-----TGSSROYSGHSFSPNIGAVNNIPVH--TLIASYNK--GFAPYGGRGY 494	
Qy	524	VATNTEVGNKHLNRSNNIELAL--GYEGDRWQYNLALYRNRFGNYIYAQTLNDGRGP 581	
Db	495	LSIDTLSSAVFNADPEYTRQYETGVKSSWLDRLSTLSAY----- 535	
Qy	582	KSIEDDEMKLVRYNQSGADFYGAEGEYFKPTPR---YRIGVSGD----- 624	
Db	536	-----QIERFN-----IRYDPKNNPIYAVSGKSRGVELSALGQI 574	
Qy	625	-----YVRGLKNLPSLPGRDAYGNRPPIAQDQNAAPRVAARLGFHLKASLTDRIAN 679	
Db	575	IPKLYLRGSLGVWQAK-----VVEDKENPDRV-----GIHL--NNTSNVTGN 615	
Qy	680	LDYR-----VFAQNKLYARET-----PGHMLN--LGANYR--NTRYGE 718	
Db	616	L-FRYTPTENLYGEIGVTGKRYGYSRKEVITLPGFARVDAMLGNHKNVNTFA- 673	
Qy	719	WNWVKADNLNLSQVVAHSSFLSDTPQMGSRFTGGVNVKF 758	
Db	674	-----AANLFNQKYWRSDSM-----PGNPGYTARVNYRF 703	
RESULT 28			
ID	AA68923	standard; Protein; 702 AA.	
AC	AA68923;		
DT	18-APR-2001	(first entry)	
DE	Neisseria meningitidis protein #22.		
KW	Meningococcus; meningitis; bacteraemia; vaccine; dsba; fhaB; fhuA;		
KW	rn15; rth; tolC.		
OS	Neisseria meningitidis.		
PN	EP1069133-A1.		
PD	17-JAN-2001.		
PF	13-JUL-1999; 99EP-0401764.		

[illegible]

```

Db 72 POTIDTLNIQKKNVGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGF 117
Qy 114 TGMADSPDHA-----IMVDTALSOQVEILRGVPTLLYSNGVAGLVVDVADGKIBKMP 168
Db 118 QADASDIYRDGVRESGQVRRSTANIERVEILKGPSSVLYGRTGGGVINMVSYANFKQS 177
Qy 169 EN--GVSG-----ELGLRSS--GNLEKLTSG--GINIGLGNFVLHT 205
Db 178 RNIGAVYGSWANRSLNMDINEVLNKNVAIRUTGEVGRANSFRSGIDSKNVWVSPSITVKL 237
Qy 206 ECLYKSGDYAVPYRNKLKLPD--SPRFANGQHRVAVLGMKRFYRRTYSDRDOYGLP-- 263
Db 238 DNLKWTGQYT--YDNVVERTPDRSP-----TKSVY-----DRFGLPY 272
Qy 264 ----AHSHEYDDCHADIIWOKSL--INIKRYLY--PHLLTEEDVDYDNFGLSCGFHD 313
Db 273 RMGFAHRNDF-----VKDKLQVW-----RSDLEY----- 296
Qy 320 HAHNGKPWIDLRNKRKYELRAEWKQFPFGFEALRVHLNRNDYHH--DEKAGDAVENFF-- 374
Db 297 -AFNDK-W-----RAQWQ-----LAHRTAAQDFDHFYAGSENGNLIKRYAW 336
Qy 375 ---NNQTONARIELR-HOPIGRKLG--SWGVOYLQK-----SSALSAT----- 412
Db 337 QOTDNKTLSSNLTLLNGDYTIIGRFENHLLTVGMDYGREHNRNPTLGFSSAFSASINPYDRASW 396
Qy 413 -SEAVKQPMLLDN--KVQHSYFF--GVEQANWD--NFTLEGGRVVEKQKASIRYDKALIDR 466
Db 397 PASGRLQPILTQNRHKADAYGIFVQNIIFSATPDLKFVLGG-----RYDKYTFNS 445
Qy 467 ENYKQPLDPAHQRQ--TARSFA--LSGNWYFTPOHKLSTASHQERLPSTQELYAHGKH 523
Db 446 ENKL-----TGSSRSYSGHSPFNIGAVMNINPVH--TLVASYNK--GFAPYGGRGY 494
Qy 524 VATNTEFVGNKHLNKRNSNIELAL--GYEGDRWQYNLALYRNFRGNVYVIAQTLDNCRGP 581
Db 495 LSINTSSAVFNADPEYTRQVETGVKSSWLDRLSTLSAVQIERFNIRY-----P 546
Qy 582 KSIEDDSEMKLVRYNQSGADFYGAEITYFKPTPRYRGVSGDYVRGLKNLPSLPQRED 641
Db 547 DBONDPYTWAVGKHRSRGVLSAIGQIIPKKL-----YLRGSLGVMAQK----- 591
Qy 642 AYNRPFTAQDONAPRYPAARLGFHLKASLTDRIDANLDYRV----- 685
Db 592 -----VVEOKENPDRV-----GIHL--NNTSNVTGNL--FFRYTPTENLYGEIGVTGTG 636
Qy 686 --FAQNKLARYETRTFGHHMLN--LGANYRNRTRYGEWNWYVKADNLNQSVYAHSSFLS 741
Db 637 KRYGYSNRNKEVTTLPGFARVADMLGNHK-----NVNITFAAANLLNQK--YWRSDAMP 689
Qy 742 DTPQGRSFTGVNVKF 758
Db 690 GAP--RTYTARVNSYF 703

RESULT 30
ID AAB68921 standard; Protein; 702 AA.
XX AAB68921;
AC AAB68921;
XX
DT 18-APR-2001 (first entry)
XX
DE Neisseria meningitidis protein #20.
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dsba; fhaB; fhuA;
XW rnl5; rth; toIC.
XX
OS Neisseria meningitidis.
XX
PN EP1069133-A1.
XX

```

```

PD 17-JAN-2001.
XX
PF 13-JUL-1999; 99EP-0401764.
XX
PR 13-JUL-1999; 99EP-0401764.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Nassif X, Tinsley C;
XX
XX WPI; 2001-082916/10.
DR N-PSDB; AAF56461.
XX
PT Immunogenic polypeptides derived from Neisseria meningitidis and the
PT nucleic acids that encode them, useful for diagnosing and vaccinating
PT against Neisseria infections e.g. bacteraemia and meningitis -
XX
XX Claim 3; Fig 20B; 240pp; English.
XX
XX The present invention provides the protein and coding sequences of
XX several genes from Neisseria meningitidis. These include the dsba, fhaB,
XX fhuA, rnl5, rth17, rth18, rth19, rth20, rth21 and toIC genes. These can
XX be used in the diagnosis and treatment of infection by the bacterium,
XX which can lead to meningitis and bacteraemia, and in vaccines to prevent
XX such infection.
XX
SQ Sequence 702 AA;
Query Match 5.1%; Score 205; DB 22; Length 702;
Best Local Similarity 20.7%; Pred. No. 2.7e-09;
Matches 178; Conservative 114; Mismatches 287; Indels 280; Gaps 50;
Qy 12 SILLINTPLLA---QAHETEQSVGLTETVTVGKSRPRATSGLLH-----TSTASDKIISG- 63
Db 12 TLIASFPPVAADTQNGEHTATLPTVSVGQSDTSVLKGVINVDAAVTRNGOLIKET 71
Qy 64 ----DTLRQKAVNLG-----DALDGVPGIHASQYGGGASAPVIRGQTRRIKVLNHHGET 114
Db 72 POTIDTLNKNKNYGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGF 117
Qy 115 GDMADSPDHA-----IMVDTALSOQVEILRGVPTLLYSNGVAGLVVDVADGKIBKMP 169
Db 118 ADASDIYRDGVRESGQVRRSTANIERVEILKGPSSVLYGRTGGGVINMVSYANFKQSR 177
Qy 170 N--GVSG-----ELGLRSS--GNLEKLTSG--GINIGLGNFVLHT 206
Db 178 NIGAVYGSWANRSLNMDINEVLNKNVAIRUTGEVGRANSFRSGIDSKNVWVSPSITVKLD 237
Qy 207 GLYKSGDYAVPYRNKLKLPD--SPRFANGQHRVAVLGMKRFYRRTYSDRDOYGLP-- 263
Db 238 NGLKWTGQYT--YDNVVERTPDRSP-----TKSVY-----DRFGLPYR 272
Qy 264 ----AHSHEY--DDCHADIIWOKSL--INIKRYLY--PHLLTEEDVDYDNFGLSCGFHD 313
Db 273 RMGFAHRNDFVKDKLQ--VWRSDEYAFNDKWRAQWQLAHTAAQDFD----- 318
Qy 314 DDDAHAAHNGKFWIDIRNKRKYELRAEWKQFPFGFEALRVHLNRNDYHHDEKAGDAVENF 373
Db 319 ---FYAGSENG-----NLIKRYN-----AWQOTDNKTLSSNFTLN--GDY-----TIGRF 358
Qy 374 FNNQTONARIELRHQ--PIGRKLSGWSGVYLGQKSSALSATSATSEAVKQPMLLDN--KVQHSY 430
Db 359 ENHLLTVGMDYSREHNRNPTLYGRSFTVPINPYDRASWPASGRL--QPILTQNRHKADSYG 416
Qy 431 FP--GVEQANWD--NFTLEGGRVVEKQKASIRYDKALIDRENYKQPLDPAHQRQ--TARS 486
Db 417 IFVQNIIFSATPDLKFVLGG-----RYDKYTFNSENKL-----TGNSRQYSGHS 459
Qy 487 FA--LSGNWYFTPOHKLSTASHQERLPSTQELYAHGKHVATNTEFVGNKHLNKRNSNI 544
Db 460 FSPNIGAVMNINPVH--TLVASYNK--GFAPYGGRGYLSITDLSAVFNADPEYTRQY 514

```


DR WPI; 1996-485781/48.
 DR N-PSDB; AAT44518.
 XX
 PT Genes encoding H. influenzae HxuC and HxuD surface-expressed
 PT protein(s) - useful in the prepn. of vaccines for children against
 PT H. influenzae infection
 XX
 PS Claim 4; Page 111-117; 188pp; English.
 XX
 CC The HxuC protein (AAW01460) of Haemophilus influenzae type b (Hib)
 CC strain DL42 is a 78 kDa outer membrane protein involved in the
 CC utilisation of low levels of free haem. It shows homology to
 CC TonB-dependent outer membrane proteins of other bacteria, and was
 CC identified from an open reading frame (see also AAT44518) located
 CC upstream of the Hib DL42 hxaA gene. Another outer membrane
 CC protein, HxuD (AAW01463), was also identified. Recombinant HxuC
 CC and HxuD proteins can be produced in transformed host cells and
 CC used to prepare vaccines for children against Hib infection;
 CC to raise diagnostic antibodies; and to prepare Hib diagnostic
 CC or therapeutic compsns.
 XX
 SQ Sequence 725 AA;
 Query Match 5.0%; Score 203; DB 17; Length 725;
 Best Local Similarity 19.8%; Pred. No. 4.4e-09;
 Matches 158; Conservative 134; Mismatches 300; Indels 206; Gaps 43;
 QY 11 LSILLINTPLLAQHETEQSGVLETVVVGKSPRATSGILHTSTASDKIISGDTLRQA 70
 DB 6 LSLAIATT--LVTAALAQSVELDSINVIATRP--SRFATPEKQSK--DSLLSKQA 57
 QY 71 VNIQDALDGVPGTHASQYGGG---ASAPVIRGTQGR-IKVNHGCTGDMADFSDDH-- 124
 DB 58 TSAVAALIEDIPNDIR--GCSRIAQKPNIRGLSDNRVQVIDGVQNFDLA-----HRG 110
 QY 125 AIMVDTALSOQVEITRGVPTLLYSSGNVAGLVADVADGKIPEKMPENGVSGLGLRLSSG- 183
 DB 111 SYFLPMSLIGIEIVIKFSSSLWCSGALGGVAMRTFNPALDLKNN---DKFGVKIRQGY 167
 QY 184 -NLEKLTSGGINT-GLGKFNVLHTEGLYKSGDYAVPRYVNLKRLPDSRPFANGQHRV 241
 DB 168 QTANLSESDVSPAANDKFDVLISGFYNNADNLRTGKGNKLN--TAVKQFGG---LAK 222
 QY 242 LGWRKRYRAT-YSDRDQYGLPAHSHYDDCHADIIWQSLINKRYLQLYPHLTEDV 300
 DB 223 FGWQINDANRVELSHRTRFQATPSN-----NEVENETNEQITDQIR 266
 QY 301 DYDNFGLSCGFHDHDDAHAAHNGKPIDLNRKRYELRAEWKQFPFGFEALRYHLNRNDY 360
 DB 267 EFHKP-----NNGSP-----PRAKPSQEEF-----YSGVKTRFGSVSY 299
 QY 361 HHDEKAGD--AVENFF--NNQTONARIELRHQPIGRKSGWGVQYLGQKSSALSATSEA 415
 DB 300 LTDQOIFDQSTVFNYLTPDNYLTHIALYNNKT-----IEKQKRVSGVKQOTKL 351
 QY 416 VKQPMLLDN--KVQHYGF-FGVEQANWDFNFTLEGGV-----RVEKQKA----- 455
 DB 352 TTRGINLRNSELISHISFVIGVDYMR-DKITERGTNNKDAQFADPDYNNANSNTTGYLI 410
 QY 456 -----SIRDY-----KALIDENYKQPLDGLGHRQTARSFALSGWYF 495
 DB 411 AHILFGEKLLSPSVRYDHTSKTVKYKDNHL-----SPATKLTWIV 455
 QY 496 TPQHKLSLTASHQE--RLPSTQELVYAGKHVAT-----NTPVGNKHLNKRERNIE 545
 DB 456 T--NWLDFTAKYNRAFRAPSQOEAFVSGSHFGTSLGNEINKE-VANPNLRPETAQKE 512
 QY 546 L-----ALGYGDRWOYNALYRPNVYIYQTLNDGRGPKSIEDSEMKLVRNQS 598
 DB 513 ITANLHFDLSFKQGDKFKEIATYFRNDVKDFINLKFNDAKTNTNASASA-----GA 564
 QY 599 GADFGVAGEIYFKPT-PRYR-----IGVSGDYVRGL---KNLPSLPQREDAYGN 645

565 GA---GANPGALLPTKSYQYQNTITNARLSGIELQAQYQTERLTLFTNYGSGTKGDKDSG- 620
 646 RPFIAQDDQNAQPRVPAARLGFHLKASLTD---RIDANLDYRVVFAQNKLYRTRT-PGH 701
 621 -----EALSNTAASKIGVGVNYALVKDKFTVGATVTHA--AQRVFPKHSVTYPSY 670
 702 HMLNLGANYRNRTRYGEW 719
 671 ILTDLRATYA--PLKGEW 686

RESULT 34
 AAW05180
 ID AAW05180 standard; Protein; 713 AA.
 XX
 AC AAW05180;
 XX
 DT 27-FEB-1998 (first entry)
 XX
 DE Neisseria gonorrhoeae strain FA19 FtpB protein.
 XX
 KW FtpB; outer-membrane protein; iron regulating; immunogen;
 KW vaccine; diagnosis; infection; Neisseria gonorrhoeae;
 KW Neisseria meningitidis.
 XX
 OS Neisseria gonorrhoeae.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..23
 FT Protein 24..713
 FT /label= FtpB
 FT /note= "70 kD major iron-regulated outer-membrane
 FT Region 32..36
 FT /label= TonB_box
 XX
 XX WO9631618-AL.
 XX
 PD 10-OCT-1996.
 XX
 XX 08-APR-1996; 96WO-US04774.
 XX
 XX 07-APR-1995; 95US-0418964.
 XX
 XX (UYNC-) UNIV NORTH CAROLINA.
 XX
 XX Beucher M, Sparling PF;
 XX
 XX WPI; 1996-465038/46.
 XX
 XX N-PSDB; AAT42127.
 XX
 PT DNA encoding Neisseria iron-regulated outer membrane protein FtpB -
 PT useful for producing vaccine to protect mammals against N.
 PT gonorrhoeae and N. meningitidis infection
 XX
 XX Claim 1; Fig 3; 68pp; English.
 XX
 PS This sequence represents a novel iron-regulated outer-membrane protein
 CC FtpB which was isolated from the gonococcal microorganism Neisseria
 CC gonorrhoeae strain FA19. This gene is also homologous to one found in
 CC Neisseria meningitidis. The FtpB protein is surface exposed and
 CC immunogenic and is predicted to be useful as a vaccine because of its
 CC surface exposure, partial antigenic conservation and susceptibility to
 CC attack by bacterial antibodies. Such a vaccine will be of use against
 CC infection in mammals by N. gonorrhoeae and N. meningitidis. The protein
 CC and its fragments will also be useful in the diagnosis of Neisseria
 CC type infections.
 XX
 SQ Sequence 713 AA;
 Query Match 5.0%; Score 202; DB 17; Length 713;
 Best Local Similarity 21.6%; Pred. No. 5.2e-09;
 Matches 187; Conservative 97; Mismatches 306; Indels 274; Gaps 46;

RESULT 35	
AAE05407	
ID	AAE05407 standard; Protein; 713 AA.
XX	
AC	AAE05407;
XX	
DT	24-SEP-2001 (first entry)
XX	
DE	N. gonorrhoeae or N. meningitidis strain FA19 FrpB protein.
XX	
XX	
KW	Fe-regulated protein B; FrpB; outer-membrane protein; vaccine; infection;

QY	384	ELRHQPIGRLKSGWVOYLGO--KSSAL-----SATSBAVKQPMLLDNKVHYS	430
DB	349	RLAEQTL--LK--YGINYRHOEIKPQAFNLSEFEIKDEKATNEKKRENEKIAKAYR	404
QY	431	FFGVEQANW-----DNFTLEGGVVRVEKQKASIRYDKALIDRENYKQPLDLAGH	480
DB	405	LTNPTKTDTGAYIEAHEIDGFTLTGGLRYDRFKVYKTHDGK-----	445
QY	481	QRTAFSFGW--YFTPOHKLSTASHQ--ERLPSTQE--LYAHGKH-----VATNTFEV	531
DB	446	--TVSSSLNPSFGVWQPREHWSFSASHNAGRPRLYDALQTHGKGIISIADGT---	500
QY	532	GNKHLNKRNNIELALGYEGDRWQYNLALYRNFPGNYIYAQTLNDG--RGPKSIEDSEM	590
DB	501	-----KAERARNTGIFNYNDGTFAAN-----GSY-FRQTIKDALANPQRHDSVAV	546
QY	591	K-----LVRYNQSADFYCAGEEYFKPTPRYRIGVSGD	624
DB	547	REAVNAGYIKHGYELGASVYRTGGLTAKVGVSHSKPRFYDTHDKLLSANPEGAQVGT	606
QY	625	YVRG---RLKNPLSLP---GR--EDAYGNRPFTIAQDDQONAPRVPAAARLGFHLKASLTD	675
DB	607	WTASLAYRFEKN--PNLEIGWRGRTYVQKAVGSIILAAQKDRDGKLENVVRQGFVN-----	659
QY	676	IDANLDYRVFAQNKLARVETRTPGHMLNLGANYRNTRYGWNWYKADNLLNQSUYA	735
DB	660	-----DVFANWKPLGKDTLN-----VNLNVN-----NVFDRKFFYP	689
QY	736	HSSFLSDT--PQMGSRFTGGVNVKF	758
DB	690	HSQRWNTNLPVGRDVRVLGVNVKF	713
RESULT 36			
ID	ABG91066	ABG91066 standard; Protein; 713 AA.	
AC	ABG91066;		
DT	29-NOV-2002 (first entry)		
DE	Neisseria gonorrhoeae outer membrane protein preFpB protein.		
KW	Gram-negative bacterial bleb; PorB; outer membrane protein;		
KW	Chlamydia trachomatis infection; Chlamydia pneumoniae infection;		
KW	Protective antigen; antibacterial; vaccine.		
OS	Neisseria gonorrhoeae.		
PN	WC200262380-A2.		
PD	15-AUG-2002.		
PP	08-FEB-2002; 2002WO-EP01356.		
PR	08-FEB-2001; 2001GB-0003169.		
PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.		
PI	Berthet FJ, Lobet Y, Poolman J, Verlant VGCL;		
XX	WPI; 2002-657510/70.		
DR	N-PSDB; ABS67385.		
XX	Novel gram-negative bacterial bleb presenting on its surface PorB outer		
PT	membrane protein from Chlamydia trachomatis or protective antigen from		
PT	Chlamydia pneumoniae, useful for preventing Chlamydia infection		
XX	Disclosure; Page 56-57; 75pp; English.		
XX	The present invention relates to a new gram-negative bacterial bleb		
CC	presenting on its surface the PorB outer membrane protein from Chlamydia		

CC	trachomatis, or a protective antigen from C. pneumoniae. The invention		
CC	is useful for preventing C. trachomatis or C. pneumoniae infection in a		
CC	host. The present amino acid sequence represents a Neisseria gonorrhoeae		
CC	protein as described in the invention.		
XX			
SQ	Sequence	713 AA;	
	Query Match	5.0%; Score 202; DB 23; Length 713;	
	Best Local Similarity	21.6%; Pred. No. 5.2e-09;	
	Matches	187; Conservative 97; Mismatches 306; Indels 274; Gaps 46;	
QY	11	LSILLINTPL-LAQAHETEQSVLETVTVVVKSRPPATSGLLHTSTASDKIISGDTLROK	69
DB	8	LSLSLSLTLAGFAHAENANVALDVTYVKGDRQ-----GSKRTINIV---TLOOK	55
QY	70	---AVNLGALDGVGPIHASOVGGASAPVI-----RGQTGRRIRK-----LNHH	111
DB	56	DESTATDMRELLKEEFSI---DFGGNGTQSQPLTLGCMGQNSVDIKVDNAYSQTLYHQ	112
QY	112	GETGDMADFPDHAIMVDTALSOQVEILRGVTVLLYSSGNVAGLVADVAGKIPKMPKPE--	169
DB	113	GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIITAKTVDAQ	153
QY	170	---NGVSGELGLRLSSG--NLEKLTSGGINIGLGNFVLHTEGLY---RKSQDVAVPP--	219
DB	154	DLKGLDKWGVKLSNGFPAGNNGASVGFQEGNF---DLFSYNRNDKDYAGKG	209
QY	220	YRN---LKRPLDSP-----RRFANGQHRAVLGWRKFPYR--RTYSRRDQGYLP	263
DB	210	FRNDGKGVYPSALDKRSYLAKEITGTFGDDHRIVLSHMKDQHRGIRTV---REEFAVS	266
QY	264	AHSEYDDCHADIIWOKSLINKRYLQYLPHLLTEEDVDYDNFGLSCGFHDDDAHAHANH	323
DB	267	EKN-----SRITIKQAPSYRE--TQSNNTLAYTKGLGFVEKLDANAYV--	310
QY	324	GKPWIDLNRKRYELRAEWKPPPGFEALRVHLNRNDYHDEKAGDAVENFFNNQOTNARI	383
DB	311	-----LEKRYK--ADDKD-----NGYAGNVKGPNHTRATSRMNFDS	348
QY	384	ELRHQPIGRLKSGWVOYLGO--KSSAL-----SATSBAVKQPMLLDNKVHYS	430
DB	349	RLAEQTL--LK--YGINYRHOEIKPQAFNLSEFEIKDEKATNEKKRENEKIAKAYR	404
QY	431	FFGVEQANW-----DNFTLEGGVVRVEKQKASIRYDKALIDRENYKQPLDLAGH	480
DB	405	LTNPTKTDTGAYIEAHEIDGFTLTGGLRYDRFKVYKTHDGK-----	445
QY	481	QRTAFSFGW--YFTPOHKLSTASHQ--ERLPSTQE--LYAHGKH-----VATNTFEV	531
DB	446	--TVSSSLNPSFGVWQPREHWSFSASHNAGRPRLYDALQTHGKGIISIADGT---	500
QY	532	GNKHLNKRNNIELALGYEGDRWQYNLALYRNFPGNYIYAQTLNDG--RGPKSIEDSEM	590
DB	501	-----KAERARNTGIFNYNDGTFAAN-----GSY-FRQTIKDALANPQRHDSVAV	546
QY	591	K-----LVRYNQSADFYCAGEEYFKPTPRYRIGVSGD	624
DB	547	REAVNAGYIKHGYELGASVYRTGGLTAKVGVSHSKPRFYDTHDKLLSANPEGAQVGT	606
QY	625	YVRG---RLKNPLSLP---GR--EDAYGNRPFTIAQDDQONAPRVPAAARLGFHLKASLTD	675
DB	607	WTASLAYRFEKN--PNLEIGWRGRTYVQKAVGSIILAAQKDRDGKLENVVRQGFVN-----	659
QY	676	IDANLDYRVFAQNKLARVETRTPGHMLNLGANYRNTRYGWNWYKADNLLNQSUYA	735
DB	660	-----DVFANWKPLGKDTLN-----VNLNVN-----NVFDRKFFYP	689
QY	736	HSSFLSDT--PQMGSRFTGGVNVKF	758
DB	690	HSQRWNTNLPVGRDVRVLGVNVKF	713
RESULT 37			

AAB68926
 ID AAB68926 standard; Protein; 703 AA.
 AC AAB68926;
 XX
 DT 18-APR-2001 (first entry)
 XX
 DE Neisseria meningitidis protein #25.
 XX
 KW Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
 KW rni5; rth; tolC.
 XX
 OS Neisseria meningitidis.
 XX
 PN EP1069133-A1.
 XX
 PD 17-JAN-2001.
 XX
 PF 13-JUL-1999; 99EP-0401764.
 XX
 PR 13-JUL-1999; 99EP-0401764.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Nassif X, Tinsley C;
 DR WPI; 2001-082916/10.
 DR N-PSDB; AAF56466.
 XX
 PT Immunogenic polypeptides derived from Neisseria meningitidis and the
 PT nucleic acids that encode them, useful for diagnosing and vaccinating
 PT against Neisseria infections e.g. bacteraemia and meningitis -
 XX
 PS Claim 3; Fig 25B; 240pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC several genes from Neisseria meningitidis. These include the dsbA, fhaB,
 CC fhuA, rni5, rth1, rth19, rth20, rth21 and tolC genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium,
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.
 XX
 SQ Sequence 703 AA;
 Query Match 5.0%; Score 200.5; DB 22; Length 703;
 Best Local Similarity 21.6%; Pred. No. 7e-09;
 Matches 186; Conservative 105; Mismatches 287; Indels 283; Gaps 50;
 QY 12 SILLINTPLLA---QAHETEQSVGLETVTVVWKSRRPRATSGLIH---TSTASDKLIISG- 63
 DB 12 TLIIASFVVAADTDNGEHYATLPTVSUVGSDTSVLKYNIDEAAVTRNGQLIKET 71
 QY 64 ----DTLR-QKAVNLG-----DALDGVPGIHASQYGGASAPVIRGQTGRRIKVLNHGGE 113
 DB 72 POTIDTLNIQKNKYNGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGF 117
 QY 114 TGMADFSPDHA-----IMVDTALSOQVEITLRGEVTLILYSSGNVAGLVADVADGKIPEKMP 168
 DB 118 QADADILYRGVRESGGVRRSTANIERVELIKPSSVLGVRINGCGVINWVKYANFKOS 177
 QY 169 EN-GVSG-----ELGLRLSS--GNLEKLTSG--GINIGLGRNFVLHT 205
 DB 178 RNIGAVTVGNARSLNMDINEVLNKNVAIRLTGEVGRANSFRSGIDSKNVMVSPSITVKL 237
 QY 206 EGLYRKSGDYAVPRYNILKLPD-SPPRFANGQRAVLGWRKFFRYRTYSDRRDQGLP- 263
 DB 238 DNGUKWTGQYT---YDNVERTPDRSP-----TRSVY-----DRFGLPY 272
 QY 264 ----AHSHEYDDCHADIWOKSLINKRYLOLYPHLLTEEDVDYDNFGLSCGFHDDDAHA 319
 DB 273 RMGFARHNDP-----VKDKLQVW-----RSDLEY----- 296

QY 320 HAHNGKFWIDLRNKRVELRAEWK-----QPPGFEALRVHLNRNDYHHDEKAGDAVEN 372
 DB 297 -AFNDR-W-----RAQWOLAHRTAAQDFDHF-----YAGSENGNLKRN 333
 QY 373 FFNNQTQNAIIELR-----HQPIGRLLG--SWGVOY-----LQOKSSALSATSEAVK 417
 DB 334 YAWQOTDNKTLSSNFTLNGDYITIGREFNHLTVGMDYSREHRNPTLGY-SRAFTASIDPYD 392
 QY 418 -----OPMLLDN--KVQHSFF--GVQANWD-NFTLEGGVVRVEKQASIRYDKA 462
 DB 393 RASWPASGRLQPILTQNRHKADSYGFVQVIFATPDLPKVLGG-----RYDKY 441
 QY 463 LIDRENVYKQPLDGLGAHQ--TARSPA--LSGNWYFTPOHKLSTLASHOERLPSTOELYA 519
 DB 442 TFENSENKL-----TGSSRQYSGHSFSPNIGAVWNINPVH--TLYASYNK--GPAPYGG 490
 QY 520 HGKHVATNTEVGNKHLNKRSSNNIELAL--GYEGDRWQYNLALYNRRFGNYIYAQTLDND 577
 DB 491 RGGYLSINTSSSAVFNADPEYTRQYETGVKSSWLDRLSTLSAYQIERFNIRYR----- 545
 QY 578 GRGPKSIEDDSEMKLVRYNOSGADFYGAEGEIVFKPTPRYRIGVSGDYVRGRLKNLPSP 637
 DB 546 ---PDEQNDPYTWAVGGKHSRGVLSAIGQIIPKXL-----YLRGSLGVWQAK- 591
 QY 638 GREDAYGNRPFIAQDDONAPRVPAARIGFHLKASLTDRIDANLDYRV----- 685
 DB 592 -----VVEDKENPDV-----GIHL--NNTSNVTGNL--FFRYTPTENLYGEIGV 632
 QY 686 -----PAQNKLYARTETPGHMLN--LGANYRRNTRYGEMWYKADNLLNQSYYAHS 737
 DB 633 TGTGKRYGYSNRKKEVTLLGPARVDAMLGWNHK-----NVNITFAANLLNOK-YMRS 685
 QY 738 SFLSDTPOMGSRSPFTGGVNVKF 758
 DB 686 DAMPGAP---RTYTARVNVYSF 703
 RESULT 38
 AAY38831
 ID AAY38831 standard; Protein; 725 AA.
 XX
 AC AAY38831;
 DT 08-OCT-1999 (first entry)
 DE Neisseria meningitidis antigen encoded by ORF23.
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
 OS Neisseria meningitidis.
 PN WO9924578-A2.
 PD 20-MAY-1999.
 PF 09-OCT-1998; 98WO-IB01665.
 PR 01-SEP-1998; 98GB-0019016.
 PR 06-NOV-1997; 97GB-0023516.
 PR 14-NOV-1997; 97GB-0024190.
 PR 18-NOV-1997; 97GB-0024386.
 PR 27-NOV-1997; 97GB-0025158.
 PR 10-DEC-1997; 97GB-0026147.
 PR 14-JAN-1998; 98GB-0000759.
 XX (CHIR-) CHIRON SPA.
 PA Grandi G, Masiagnani V, Pizza M, Rappuoli R, Scarlato V;
 PI WPI; 1999-327407/27.
 DR N-PSDB; AAZ12259.
 XX

PT Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for
 PT diagnosis, treatment and prevention of infection
 XX
 PS Claim 4; Page 379-380; 524pp; English.
 XX
 CC Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*
 CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open
 CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
 CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
 CC infections, such as meningitis, septicaemia and gonorrhea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.
 XX
 SQ Sequence 725 AA;
 Query Match 5.0%; Score 200; DB 20; Length 725;
 Best Local Similarity 21.7%; Pred. No. 8.2e-09;
 Matches 178; Conservative 105; Mismatches 294; Indels 244; Gaps 43;
 QY 19 PLLAQAH-----ETEQSVGLTETVTVVVKSRPRATSGLLHTST-----AS 57
 DB 15 PVYQAADVSVDDPKQESTELTITVADRATSSNDGYTVSGTHPLGLEPMTLEIPQS 74
 QY 58 DKISGDTLR-QKAVNLGDALDGVPGHISQYG---GGASAPVIRGOTGRRYKVLNHGCE 113
 DB 75 VSVITSQMDQDKIKTLDRALLQATGTSRQIYSDRAGYNLFARGS-----KIANQIN 129
 QY 114 TGMADFPDPAHIMVDITALSQVEILRGPVTLVYSSGNVAGLVADVADGKIPKMPENGYS 173
 DB 130 GIPVADALADTG-NANTAAVERVEVVRGVAGLLDGTGEPATVNLVKRLTRK-PLFEVR 187
 QY 174 GELGLRLSSNLEKLTSGGINIGLGNFVLHTEGLRKSDYAVPRYNLKRLPDPSPRF 233
 DB 188 AEAGNRKHFG-LDADVSGSLN-----TEGTLR-----GRIVST---F 220
 QY 234 ANGOHRAVLGRKFRFRTYSDRRDQYGLPAHSHEVD-----DCHADITWQSLINKRYL 288
 DB 221 GRGD-----SNR-----RRERSDAELXGI---LEYDIAPTQTRVHAGMDYQQA----- 260
 QY 289 QLYPHLLTEEDV-----YDNLGLSCGHHDDAHAAHNKG-----PWIDLR-NK 333
 DB 261 -----KETADAPLSYAVYDVSQYATAFGPKDNPATNWSRHRALNLFAGIEHRFNQ 312
 QY 334 RYELRAEW-----KOPFPGFEALRVHLNRNDVHDEKAGDAVENFF---NNOTQARIE 384
 DB 313 DWKLKABDYTRFRFPYGVAGVLSI-----DHNTAATDLIPGVHADPTHSASVS 365
 QY 385 L--RHQPIGR-----LKGSGVQYLGQKSSALSATSEAV-----KQMLLDNKV 426
 DB 366 LIGKYLFGREHDLIAGINGYKASNKYGIERSIIPNAIPNAYEFSRTGAYPQSPASQTI 425
 QY 427 QHY-----SFFGVEQANWDFTEGGVVRVEKQASIRYDKALIDRENYKQPLDGA 479
 DB 426 POYTRQIGYLAIRPADNLSLILGGRYTRTG-SYDSR-----TGMYTVA 476
 QY 480 HQTARS---FALSGNWFYTPQHKLSTASHQERLPSTQELYARHGKHAIVATNTEVGNKHL 536
 DB 477 NRFTPTGIYFDLGN-----LSLYGVSYSFLFVPSQKDEHGSYLKPVIT----- 520
 QY 537 NKESNNIELALGYED-----RWQYNLALYRNRGNVYIYAQTLNDGGRKSIEDDSEML 592
 DB 521 -----GNLLE---AGIKGEWLEGRNASAAVYRARNLNATAA-----GRDP----- 559
 QY 593 VRYNSQADFY-----GAEGEYFKPTPRYRI---GVSGDYVR---GRKLNPLSLPG 638
 DB 560 -----SGNTYRAANQAKTHGEIVGGRITPEQIQAGYSQSKTRDQSGRLNPDVSPE 614
 QY 639 RE-----DAYGNRPPIAQDDQNAVRPAARLGFHLKASLTDRIDANLDYRVFQNKIARY 694
 DB 615 RSFKLFTAY-----HFAPEAPS---GWTIGAGVRWQSETHPTPATIRIPNAAKA 661
 QY 695 ---ETRTPGHMLNLGNARYRNTRYGEMWVYKADNLNQ 731

DB 662 RAADNSRQKAVAVADIMARYRFPNRA---ELSLNVDNLFNK 699
 RESULT 39
 AAE05851
 ID AAE05851 standard; Protein; 687 AA.
 AC AAE05851;
 DT 24-SEP-2001 (first entry)
 DE *Pseudomonas stutzeri* open reading frame-K (ORF-K) protein.
 KW Open reading frame-K; ORF-K; pyridine-2,6-bis (thiocarboxylate);
 KW PDC; environmental remediation; phytoremediation; bioaccumulation;
 KW water purification; solution mining mobilisation; immobilisation;
 KW detoxification; redox state modifier; metal ion reactivity;
 KW carbon tetrachloride; metal.
 XX
 OS *Pseudomonas stutzeri*.
 XX WO200153309-A1.
 XX 26-JUL-2001.
 XX 19-JAN-2001; 2001WO-US02386.
 XX 20-JAN-2000; 2000US-0177251.
 XX (IDAH-) IDAHO RES FOUND INC.
 XX (PASZ/) PASZCZYNSKI A.
 XX (SEBA/) SEBAT J L.
 XX Paszczynski A, Sebat JL, Lewis TA, Crawford RL, Cortese MS;
 DR WPI; 2001-465361/50.
 DR N-PSDB; AAD11171.
 XX New isolated nucleic acids, useful for producing enzymes required to
 PT produce pyridine-2,6-bis (thiocarboxylate), especially useful for
 PT reducing the amount of metal or carbon tetrachloride in a substrate,
 PT e.g. soil or water
 XX
 PS Claim 33; Page 95-98; 172pp; English.
 CC The present sequence is a *Pseudomonas stutzeri* open reading frame-K
 CC (ORF-K) protein. The *Pseudomonas stutzeri* genome includes ORFs that
 CC encode enzymes required for biosynthesis of pyridine-2,6-bis
 CC (thiocarboxylate) (PDC). The ORFs encoding PDC are especially useful
 CC in environmental remediation methods, e.g. phytoremediation,
 CC bioaccumulation, water purification, waste water purification, solution
 CC mining mobilisation, immobilisation, detoxification, redox state
 CC modifier or modification of metal ion reactivity. In particular, the
 CC ORFs are useful for degrading carbon tetrachloride and removing metals
 CC from substrates, e.g. soil or water.
 XX
 SQ Sequence 687 AA;
 Query Match 4.9%; Score 198.5; DB 22; Length 687;
 Best Local Similarity 21.4%; Pred. No. 1e-08;
 Matches 170; Conservative 111; Mismatches 286; Indels 229; Gaps 43;
 QY 7 KPVLISILLINT-----PLLAQA-----HTEQSVGLTETVTVVVKSRPRATSGLLHTSTAD 58
 DB 21 QPMWATALICAFVPGPQLAFAPGSAASPDSTTLPEITVTAERPE---LERVPASV 76
 QY 59 KIISGDTLRQKAVNLGDALDGVPGHISQYG---GGASAPVIRGOTGRRYKVLNHGCE 116
 DB 77 AVIDGDWAEQSGITSLKQLEGRIPGLSPFQPGQAMNSPVMRGL-----TAN 123
 QY 117 MADFPDPAHIMVD---TALSO-----QVETLRGPVTLVYSSGNVAGLVADVADGKI 163

Db 124 PNFSSSTLLIADGVPTLTQAFSGSGLDLDRIVIRGPOSTLYGRNAEAGV--IAHSL 181
 Qy 164 P-EKMPENGSGELGLR-----LSSGNLEKLTSGGI--NIGLGNFVLHTEGLYRKS 212
 Db 182 PMDATPRASVAEAGSRNKRVMRFALSOPLVEERLYGSVGNSSQDGFIDNTH-TGHKA 240
 Qy 213 GDYAVPRYRNKRLPDSPPRFANGCHRAVLGWRKRFYRTYSDRRDQYGLPAHSHEYDDC 272
 Db 241 DD-----REQNL-NLGLRPAAGTAVV--NRYAHQEVYDDGASLWSPGAPRK-RVA 289
 Qy 273 HADIWQSLNKRILYQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAHANGKRWIDLRN 332
 Db 290 SGTSPSWNS-----EGQTLSPNVQHFASGLRLHSVTAWNEFD 328
 Qy 333 KYELRAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFNQTNQARIELRHQIGR 392
 Db 329 -----RIQDQTFMPADVL--HVGED--HH-----LRTLSQEFVVE-----GQ 362
 Qy 393 L-KGSW--GVQVLGQKSSALSATSEAVKQPMMLLNKVQHYFFGVQANWD-----NFTL 444
 Db 363 LGEASWLAGV-YADRSNDLHSTSKTMWG--LSDIRAQDQSDTAALFTHMNVPLSADWSI 419
 Qy 445 EGGVEVEKQASIRYDKALIDRENYYKQPLDGLGAHROTARSFALSGWNYFTPPQHKLST 504
 Db 420 DAGARVERNEQLR-PQATSEKQWTHVSPRLALHQ-----IT 458
 Qy 505 ASHQBELSPSTOELYAHGKHVATNTPEVGNKHLNKRNSNIELALGYEG-----DRWQYNLA 560
 Db 459 ANHQWYVASASRGVGTGGFNLAFTL-----GYLPYDTEKNWSYETGLKWLIDKRIYSLA 514
 Qy 561 LYRNFNGYIYAQTLDNGRGPKSIEDDSEML-----VRYNOSG--ADPYGAGEIYKPT 614
 Db 515 AY-----LMD-----IDMQVQMPVGVWMIITSAATATSKGLELDV----- 551
 Qy 615 PYRIGVSGDYVRG---BLKNLPS-----LPREDAYGNRPFFIAQD-----DON 655
 Db 552 -----DYLLGGWQLKGGLANWTRFDHFRDGEADYDGNQNPAPDLTGLHGYRD 602
 Qy 656 APRVPAARLGFHLKSLDRIDANLDYRVFAQNKLAARYETETPCHHMLNLANRYNRTR 715
 Db 603 APE-----GWAQASVGTSSKVIYLDAAANGYERN-----GYGLNVLVAGYQ--- 643
 Qy 716 YGEMNMYKADNLLNQ 731
 Db 644 -GNWEIAVADNATDQ 658
 RESULT 40
 AAY34501
 ID AAY34501 standard; Protein; 708 AA.
 AC AAY34501;
 XX
 XX 20-MAR-2003 (updated)
 DT 25-AUG-1999 (first entry)
 DE
 DE Porphyromonas gingivalis protein PG40.
 XX
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 XX
 OS Porphyromonas gingivalis.
 XX
 PN W09929870-A1.
 XX
 PD 17-JUN-1999.
 XX
 XX 10-DEC-1998; 98WO-AU01023.
 PF
 XX 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 PR 30-JUL-1998; 98AU-0004963.
 PR 04-AUG-1998; 98AU-0005028.
 XX (CSLC-) CSL LTD.
 XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 XX WPI; 1999-385613/32.
 DR N-PSDB; AAX91719.
 DR Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 PT
 XX
 PS
 PS Claim 1; Page 482-483; 58pp; English.
 XX
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC (Updated on 20-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 708 AA;
 Query Match 4.9%; Score 198.5; DB 20; Length 708;
 Best Local Similarity 20.1%; Pred. No. 1.1e-08;
 Matches 163; Conservative 103; Mismatches 302; Indels 241; Gaps 35;
 Qy 15 LINTPLAQAHETEQSV-----GLETVVVGVKSRPRATSGLLHTSTASDKIISGDTL 66
 Db 8 LLSIGISAQAFKTDNVTDSLRVHNLQTVTVYS-----TRTAVPLKKIPAKMELISSRNI 63
 Qy 67 RQKAV-NLGDALDGVPIHASOYGGASAPVIRG--QTGRRIKVLNHHGETGMADFSFD 123
 Db 64 KQSGFNWMTDLTKTQSSLDVIQYPGFSSNIGIRGPKSKYTVLVNGIPAG-----TD 117
 Qy 124 HALMVDTALSQQVEILRGVPTLLYSSGNVAGLVVDVADGKIPEKMPEN-----GVSG 174
 Db 118 NISTLNTSNIEQIEILKGPFSIYGTNMGVVVNIITHKSKDKIHGNVSLFGGSYQTMAG 177
 Qy 175 ELGLRLSSGNLEKLTSGGINIGLKG-----NF-----VLHTEGLYRKS 212
 Db 178 SFNL--GGRFEDIFSDISLGLDKQNDKYKTSNNFLSKLEEAIVDVNATKMKMG 234
 Qy 213 GDYAVPRYRNKRLPDSPPRFANGCHRAVLG-----WRKRFYRTYSDRRDQYGLP 263
 Db 235 SDYTVATGRLRFGIDFTPEWSNLNYQNVLGDAIPVGGSIW-----GVYGESKKNLARS 288
 Qy 264 AHSHEY---DDCHADIIWQSLNKLKEYLQYPHLLTEEDVDYD--PGLSCGFHDD----- 314
 Db 289 STSFELLGKHGCHT-----LQSPYFNTEKSENYNADPTGFIYKSDYTY 335
 Qy 315 -----DDAHAHANGKFWIDLRN-----KRYELRAEWKQFP-PGF----- 348
 Db 336 GALLQDKISFGGQIVLGVDSRNWMTESERFEQAGVNTKPYNPGYATNIGLFGQANFYL 395
 Qy 349 --BALRVHLN-RNDY--HHDEKAGDAVENFNQTNQARIELRHQIPGRKLGWGVQYILQ 404
 Db 396 LNDALISISAGARADFMFELKANEXYLNNEAKQETHNV-----INPNVGIKYEV 444
 Qy 405 KSSALSAT--SEAVKQPMMLDNKVQHYFFGVQANWD-----NFTLEGVVRVEKOKASIR 458
 Db 445 KGLTAHCTFGSAPFADPAQKAGQYVGPFGTIGNIPDLKPEKSMWTDFIGISNARCGIO 504

QY 459 YDKALIDRENYKQPLDPLGAHRQTARSFALSGNWYFTPOHKLSTASHOERLPSTQELY 518
 Db 505 ADVTLT-----YFTHDKOLILSSPD-----Y 526
 QY 519 AHGKHVAINTPEVGNKHLNKNERSNNIELALGYEGDRWOYNLALYRNFGNYIYAQ-TLND 577
 Db 527 ANNITTYINA-----DKARMSGIEALLSYDFG-----SLFANKFSLRAFANATIML 572
 QY 578 GRGPKSIEDD---SEMKLVRYNQ--SGADFYGARGEIYFKPTPRYRIGVSGDYVRGLKN 632
 Db 573 NSEMKSTDPAPWSEMYVRKONITFGIEYRKEG-----LEVMLNGRFMRRIE- 622
 QY 633 LPSLFGREDAYGNRPFFIAQDDQ-----NAPRVPA-----ARLGFHLKASLTD 674
 Db 623 -----QNWYAYYPEVRPELQOLLAABEPELAAQGLLEHPQAMVFNASAYYHMKYLT 675
 QY 675 RIDANLDYRVRFAQNKLARYETRTFGHHM 703
 Db 676 GVNLN-----NILDELYTEKDGYYHM 695

Search completed: November 14, 2003, 10:58:40
 Job time : 55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:56:57 ; Search time 21 Seconds

(without alignments)
3471.231 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLKPIVLSILLINTEL.....FLSDTFQMGSRSTGGVNVKF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.*

1: Pirl.*

2: Pirl2.*

3: Pirl3.*

4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3845	95.3	764	2 D81883	probable outer-mem
2	694.5	17.2	681	2 E82812	outer membrane hem
3	679	16.8	687	2 D83547	hypothetical prote
4	636	15.8	687	2 E87275	TonB-dependent rec
5	290	7.2	713	2 F82506	probable TonB syst
6	264.5	6.6	715	2 H90977	hypothetical prote
7	264.5	6.6	719	2 F85824	hypothetical prote
8	246	6.1	721	2 E83237	probable TonB-depe
9	243	6.0	680	2 H70347	outer membrane pro
10	231.5	5.7	764	2 H83055	probable outer mem
11	231	5.7	652	2 D82317	iron-regulated out
12	228	5.6	665	2 AD0160	probable outer mem
13	228	5.6	744	2 B64049	outer membrane pro
14	227	5.6	652	2 S25265	outer membrane pro
15	224	5.6	723	2 F83173	outer membrane pro
16	221	5.5	723	2 C84058	outer membrane pro
17	214.5	5.3	720	2 S81133	Fe-regulated prote
18	213.5	5.3	714	2 F81962	probable iron-regu
19	212.5	5.3	703	2 D81215	TonB-dependent rec
20	212.5	5.3	714	2 C81018	iron-regulated out
21	211	5.2	673	2 A56148	pesticin receptor
22	211	5.2	673	2 AF0232	pesticin/yersiniab
23	210	5.2	673	2 S60142	outer membrane pro
24	208.5	5.2	703	2 E81792	probable TonB-depe
25	208	5.2	710	2 E83360	conserved hypothet
26	205.5	5.1	676	2 T12069	hemin receptor - y
27	205.5	5.1	676	2 AG0035	hemin receptor pre
28	203	5.0	725	2 A57148	outer membrane pro
29	203	5.0	802	2 AC2075	outer membrane hem

Fe-regulated prote
ferrienterobactin
TonB-dependent rec
probable TonB-depe
probable ferric si
probable vitamin B
heme utilization/t
outer membrane hem
metal chelate oute
colicin I receptor
probable TonB-depe
hypothetical prote
hypothetical prote
colicin I receptor
vitamin B12 recept
hemin receptor pre

30 202 5.0 713 2 A56268
31 200.5 5.0 751 2 AC0574
32 200 5.0 725 2 H81030
33 199 4.9 616 2 E83485
34 192 4.8 725 2 D81976
35 191 4.7 625 2 AD0476
36 188 4.7 660 2 D91176
37 188 4.7 660 2 E86022
38 187.5 4.6 599 2 AC3334
39 187.5 4.6 663 1 QREIC
40 187.5 4.6 690 2 A80124
41 185 4.6 614 2 A88241
42 185 4.6 614 2 F86088
43 184.5 4.6 659 2 G91009
44 182.5 4.5 614 2 AB0935
45 181 4.5 686 2 S28042

ALIGNMENTS

RESULT 1

D81883
probable outer-membrane receptor protein NMA1161 [imported] - Neisseria meningitidis (s
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: D81883
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: D81883
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-764 <PAR>
A;Cross-references: GB:AL162755; GB:AL157959; NID:G7379742; PIDN:CA884423.1; PID:G73798
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA1161

Query Match 95.3%; Score 3845; DB 2; Length 764;
Best Local Similarity 95.9%; Pred. No. 6.2e-266;
Matches 730; Conservative 7; Mismatches 18; Indels 6; Gaps 3;

Qy 1 MAQTTLKPIVLSILLINTELPLAQAHTEQSVGLTVTVVVGKSRPRATSGLLHTSTASDKI 60
Db 7 MAQTTLKPIVLSILLINTELPLAQAHTEQSVGLTVTVVVGKSRPRATSGLLHTSTASDKI 66
Qy 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIRKVLNHHGETGMDAPF 120
Db 67 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGOTGRRIRKVLNHHGETGMDAPF 126
Qy 121 SPDHAIMVDTALSSQVEILLRGPVTLTYSSGNVAGLVADVADGKIPEKMPENGVSGLGLRL 180
Db 127 SPDHAIMVDTALSSQVEILLRGPVTLTYSSGNVAGLVADVADGKIPEKMPENGVSGLGLRL 186
Qy 181 SSGNLEKLTSGGINIGLKNVLTGLYKSGDYAVPYRNILKLPDSPRFRANCOHPA 240
Db 187 SSGNLEKLTSGGINIGLKNVLTGLYKSGDYAVPYRNILKLPDSPRFRANCOHPA 243
Qy 241 V-LGW--RKRFRVRRYTSRRDDQVGLPAHSHEDDCHADIIWQKSLINKRYLQYPLHLTE 297
Db 244 IGLSWGKEGFGIAAYSDRRDQVGLPAHSHEDDCHADIIWQKSLINKRYLQYPLHLTE 303
Qy 298 EDVDYDNPGLSCGFHDDDDAHAAHNGKWDILRNKRYELRAEWKQPPFGPEALRVHLNR 357
Db 304 EDYDYNPGLSCGFHDDDDAHAAHNGKWDILRNKRYELRAEWKQPPFGPEALRVHLNR 363
Qy 358 NDYHDEKAGDAVENFNNTONARIELRHOPIGRLKSGVQYLGQKSSALSATSEAVK 417
Db 364 NDRHDEKAGDAVENFNNTONARIELRHOPIGRLKSGVQYLGQKSSALSATSEAVK 423


```

QY 418 QPMLLDNKVQHYSPFGVEQANWNTLEGGVYRVEKASIRYDKALIDRENYKQPLPDL 477
DB 424 QPMLLDNKVQHYSPFGVEQANWNTLEGGVYRVEKASIRYDKALIDRENYKQPLPDL 483
QY 478 GAHROTARFALSGNWNFFPQHKLSLTASHQBELPSTOLYAHGKHVATNTFEVGNKHLN 537
DB 484 GAHROTARFALSGNWNFFPQHKLSLTASHQBELPSTOLYAHGKHVATNTFEVGNKHLN 543
QY 538 KERSNNIELALGVEGRDQWYNALYRNRFGNIIYAQTLDNGRGPKEIDDDSEMKLVRYNQ 597
DB 544 KERSNNIELALGVEGRDQWYNALYRNRFGNIIYAQTLDNGRGPKEIDDDSEMKLVRYNQ 603
QY 598 SGADFYGAGEIYFKETPRYRIGVSDYVRGLKMLPSLPGREDAYGNRPFTAQDQNP 657
DB 604 SGADFYGAGEIYFKETPRYRIGVSDYVRGLKMLPSLPGREDAYGNRPFTAQDQNP 663
QY 658 RVPAAELGPHLKASLTDRIDANLDYRVPFAONKLAETRTTPGHMLNLGANYRNTRYG 717
DB 664 RVPAAELGPHLKASLTDRIDANLDYRVPFAONKLAETRTTPGHMLNLGANYRNTRYG 723
QY 718 ENNNYVKADNLLNQSVYAHSSFLSDTPQMRSGFTGGVNVKF 758
DB 724 ENNNYVKADNLLNQSVYAHSSFLSDTPQMRSGFTGGVNVKF 764

RESULT 2
E82812
outer membrane hemin receptor XF0384 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 23-Dec-2002
C:Accession: E82812
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82812
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-681 <SIM>
A:Cross-references: GB:AE003890; GB:AE003849; NID:g9105215; PIDN:AAF83194.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorfi, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Mitacca, B.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaak
M.; Tshako, M.H.; Vallada, H.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0384
C:Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to
Query Match 17.2%; Score 694.5; DB 2; Length 681;
Best Local Similarity 27.4%; Pred. No. 2.3e-41;
Matches 220; Conservative 115; Mismatches 294; Indels 173; Gaps 22;

QY 1 MAQTTLPVLSILLINTPLLAQA-----HETQSVGLETVTVVG----- 40
DB 9 LPRVFLVAVSSLL--TPALAMADVCPPTDSHFKTTEVTHLKDLDVAVVTAIPLRD 66
QY 41 ----KSRPRATSGLLHTSTASDKTIGSDTLRQ-KAVNLGDALDGVFGIHASQYGGGASAP 95
DB 67 SYSDLRSFVA-----LLAGERLDEVRASSLIGETVAVLPGVQSSNPGVGWRP 113
QY 96 VIRGQTGRIRIKVNLHGETGMDADFSFDHAIWDTALSQQVEILRGVPTVILYSSGNVAGL 155

```

```

DB 114 IIRGLDPRVAVLNLNGLSQDVSIVSQDSHPAVEPFLANQIEVLKGPSILLYSGAIGGV 173
QY 156 VDVAADKIPKMPENGVSGBELGLRLSSGNLEKUTS--GGINIGLGKNFVLHTEGLYKSGD 214
DB 174 VNIVDGRIAE-APVGGFNGRAEMRLDGDGKHGNTNMFRIADAGNGSALSADGVYREXND 232
QY 215 YAVPRYENLKLDPSPRRFANGQHRVAVLWGRKFEYRTYSDDRDQYGLPAHSHEYDDCHA 274
DB 233 YDFPKGQVNSPFDT-----KSGMNGASFGSDGFGFGLSVARFHDYSG----- 275
QY 275 DIIWQKSLINKRYLQYLPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHAKHAKPWIDLENKR 334
DB 276 -----NPG-----EPGDPVAGDRGSMRLRHQDR 298
QY 335 YELRAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFNNQTNQARIELRHQPTGLRK 394
DB 299 YDLKAGLTDPWGBASALRFLSGHTQVDHIEFEGNEVGTTFGKRASEGRVEASFAFGGWR 358
QY 395 GSWGV-----QYLGKSSALSATSEAVKQPMLLDNKVQHYFFGVGEQANWNTFLEGG 447
DB 359 TAFVGQGDSTFQALGEESFVPKTNKSI-----GVFGLAHTNTFGLFQAEPG 405
QY 448 VRVEKQKASTRYDKALIDRENYKQPLDLCARHQTARSPALSGNWNFFPQHKLSLTASH 507
DB 406 ARGDQ-----VKYDNDNGVTRNYH-----PGSLAFSGDLALSQWELTLNVDH 448
QY 508 QERLPSTOEIVAHGKHVATNTFEVGNKHLNKRSSNNIELALGYEGDRQWYNALYRNRFG 567
DB 449 AERAPVEEELFAKQPHIATLAYEVRADLKKERANQALGLVFNDSWDKAVTYYSRYG 508
QY 568 NYIYAQTLDNGRGPKEIDSEMKLV--RYNQSGADFYGAGEIYF-----KPTPRYRIGV 621
DB 509 NFIIY---LVD--TGSTWFWDDQDRDLPVRQWSQANAI FHIIEGEATFHLAKNTSGSMWDLRV 564
QY 622 SGDVRGLKMLPSLPGREDAYGNRPFAODDQNAVRVPAARLGFHLKASLTDRIDA--- 678
DB 565 PGDTVSGRLKN-----GNLPRIVPARYGAELRWE-----DAGWR 599
QY 679 -NLDIYRVFAONKLAETRTTPGHMLNLGANYRNTRYGEMWYVKADNLLNQSVYAH 737
DB 600 TSLSAKRYEKQKQVAVNETTAGYTMVDLAHLAYHIDVDSIAWEVFFPDGNNLTNRDARVHT 659
QY 738 SFLSDTPQM-CRSFTGGVNVKF 758
DB 660 SFLKDDVNLGRNTYTAGLRMPF 681

RESULT 3
D83547
hypothetical protein PA0781 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83547
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; E
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lir
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: A82950; MUID:2043737; PMID:10984043
A:Accession: D83547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE004513; GB:AB004091; NID:g9946669; PIDN:ARG04170.1; GSPDB:GN00
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0781

Query Match 16.8%; Score 679; DB 2; Length 687;
Best Local Similarity 27.4%; Pred. No. 2.9e-40;
Matches 212; Conservative 121; Mismatches 319; Indels 122; Gaps 25;

```

```

Qy 6 LKPIVLSILLINTPLL-AQAHETEVSGLTETVTVGKSRPRATSGLLHTSTASDKIISGO 64
Db 15 LTPALACLVSGEGLGADGPSELPSQVITANPLGNESPATPS-----SVLEGD 64

Qy 65 ---TLQKAVNIGDALDGVPGHSGGASAPVIRGOTGRRIKVLNHHGTGMDAPSP 122
Db 65 ELTLRQKG-SLGETLNGLPVGSSTVFGPGASRPVIRGMDGRIIRLLRNGVGLDASSLSY 123

Qy 123 DHAIWVDTALSOQVEILRGPTVTLVSSGNVAGLVADVADGKIPEKMPENGVSGLGLRLSS 182
Db 124 DHAVPEDNSVERLEVRGPAALYCGNAIGGVNVSFNRIPSE-PVDGIHSGELRYGG 182

Qy 183 GNLEKLTGGINIGLKNPVLHTEGLYKSGDYAVPRYRNKRLPDSPPR-ANGQHRA 240
Db 183 ADTTSRSGALEAGDG-NFALHVDAAAREFNDVIRPGVAHSSR---QRQIDGDTGKHV 237

Qy 241 VLGRKRFYRRYTSRRDOYGLPAHSHEVDDCHADI IWQSLINKRYLYLPHLLTREDV 300
Db 238 ---QNSDQDGGVAGVSYHWEHYAGLSY-----SGYDS 269

Qy 301 DYDNFGLSCGFHDDDAHAHAHNGKRWIDLRNKRVELRAEWKQPPFGFEALRVHLNRNDY 360
Db 270 NYGSPA-----EDDVR-----LKMQODRYAFASEIRDLRGPTSLKLDAAATKY 313

Qy 361 HHDEKAGDAVENFNNOTQNAIRIELRHOPIGRLKSGWGVYLGQKSSALSATSEAVKQPM 420
Db 314 EHKEIEDGETGTTFKNEGYEGRIARHRLPLGLNGVGAQFANSRFSALG-----EEAF 367

Qy 421 LLDNKKVQHYSPFGVEQANW---DNFTLEGGVVERKOKASIRYDKALIDRENYKQPLDL 477
Db 368 VPHETETDSAAALFALEE-WKLSRDLQSGARLE---HTRVDPDAKGNRFAEND--- 417

Qy 478 GAHRQTARSAFALSGWYFTPOHKLISLTASHQBELPSTQELYAHGKHVATNTFEVGNKHLN 537
Db 418 GSCQFTTGLSGLSGAVYKLTPIWLSAATLSYTERAPTVELYANGPHAATGTVEVGDAAD 477

Qy 538 KERSNNIELALGYEGDRWQYNALYRNRFNGYIYAQTINDGRGPK-----SIEDDSBMK 591
Db 478 KEKAVSTDALAFDNGVHKSGVGYFRSPNVI---GLLASGRHREGEVVAAGDDEALP 535

Qy 592 LVRYNQSADPYGAGE-----YFKPTPRYRIGVSGDYVRGRLKMLPSLPGREDAYGNRP 647
Db 536 EYLVSGRADFYGVEAQRDHIHLESYPYGNFDLESGDYTRAKNKD----- 580

Qy 648 FIAQDDQNAIPRYPAARLGHKLKASLTDRIDANLDYRVFAQNKLAARYETRTPGHMLNIG 707
Db 581 ---TGEPLRIAPLRDNTALIWEL---QQWQARVDVVEHAASQHRVPEEBELSTDGY---TTIG 633

Qy 708 ANYRNRTRYGEWNW---YVKADNLLNQSVYVAHSSFLSD-TPQMGRSFTGGVNVYKF 758
Db 634 ASLGYNFDLGEGRWLAFAVKGTLNLTQTVRYASSILRDRVPAAGRGIEAGVKVAF 687

```

RESULT 4

```

E87275
TonB-dependent receptor [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87275
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
n, J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87275
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE005673; NID:g13421339; PIDN:AAK22201.1; GSPDB:GN00148
A:Genetics:
A:Gene: CC0214

```

RESULT 5

```

F82506
probable TonB system receptor VCA0064 [imported] - Vibrio cholerae (strain N16961 seros
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82506
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
J., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: F82506
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-713 <HEI>

```

```

Query Match 15.8%; Score 636; DB 2; Length 687;
Best Local Similarity 26.5%; Pred. No. 3.4e-37;
Matches 206; Conservative 107; Mismatches 397; Indels 168; Gaps 23;

Qy 22 AQAHETEVSGLTETVTVGKSRPRATSGLLHTSTASDKIISGDTLRQKA-VNLGDALDGV 80
Db 37 ASAQDAEVSKVITTAAPYGV-----ADALTASVAVLDRTDLDLAAPKGLGDALAGL 88

Qy 81 PGHASOYGGASAPVIRGOTGRRIKVLNHHGTGMDADFSPOHAIMVDTALSOQVEILR 140
Db 89 PGVRSITTFGAGASRPVVRGLAGRPVQVLTNGVQIDASALSPOHQVATDPGEAERIEVL 148

Qy 141 GPTVLYSSGNVAGLVADVADGKIPEKMPENGVSGLGLRLSSGNLEKLTSGGINIGLKN 200
Db 149 GPAALYGGSAIGGVNIIIDRISTQOPIIDGMSGRLLASRGTDGDSYALSAGVAHATVGP- 207

Qy 201 FVLHTEGLYKSGDYAVPRYRNKRLPDSPPRFA-----NGOHR-----AV 241
Db 208 MVLTDALKRESKDYKIPY-----PESARQALAEGETAEGGAGRLSENSAVDLETFGAG 261

Qy 242 LGM---RKRFRYRRTYSDRRDOYGLPAHSHEVDDCHADI IWQSLINKRYLYLPHLLTEED 299
Db 262 LSYVVDKGFVMSIKRTDSTYGVGHAHE----- 290

Qy 300 VDYNPGLSCGFHDDDAHAHAHNGKRWIDLRNKRVELRAEWKQPPFGFEALRVHLNRND 359
Db 291 -----HEGEAEAGHBESAVTIGLKOTRIDLRGEYDADLGPFKAVRFGSGHAD 338

Qy 360 YHDEKAGDAVENFNNOTQNAIRIELRHOPIGRLKSGWGVYLGQKSSALSATSEAVKQPM 419
Db 339 YTHTEPEGDAVGTKFTSDGYEGRLELVQTERGGWKGAVGVQALRRNRPDAIGDEAYVVK-- 396

Qy 420 MLDNKKVQHYSPFGVEQANWNTFLEGGVVERKOKASIRYDKALIDRENYKQPLDLGA 479
Db 397 ---TKITEGCAFTQORLELDGYEGGLRIDTRELDSLKGKA---DFTN----- 439

Qy 480 HRQTARSAFALSGN---WYFTPOHKLISLTASHQBELPSTQELYAHGKHVATNTFEVGNKH 535
Db 440 -----LSGSVGAFWRPTTSEFVGLALSRSRAPTESELFAGGPHAATRGFEIGDAD 490

Qy 536 LNKERSNNIELALGYEGDRWQYNALYRNRFNGYIYAQTINDGRGPKSIEDDSMKLVRY 595
Db 491 LKBTATSLKATLHYGGERVSGDLHLIYARYDGFIDLRPTGD-----BEDG-LAVRY 542

Qy 596 NOSCADFVAGEITVFKPTPRYRIGVSG-----DYVRGLKMLPSLPGREDAYGN 645
Db 543 VQDAEPRGFEAEI-----AYRVTMGQSRVNLHAGADPVRG----- 579

Qy 646 RPTIAQDDQNAIPRYPAARLGHKLKASLTDRIDANLDYRVFAQNKLAARYETRTPGHMLN 703
Db 580 -----SSDLGPPARIIPPYSVS--LKASYQAPMWSGDVEVRRTGQQRVAQWELPTDGYTV 632

Qy 704 LALCANRY--RNTRYGEWNWYVKADNLLNQSVYVAHSSFLSD-TPQMGRSFTGGVNVYKF 758
Db 633 LNAATLAWKPAGNARV---RULFDGRNLTNEEVREHVSFLKDIAPSPGQVRAGIALRF 687

```

A; Cross-references: GB:AE004349; GB:AE003853; MID:99657434; PIDN:AAF95978.1; GSPDB:GN001
A; Experimental source: serogroup O1; strain N15961; biotype El Tor
C; Genetics:
A; Gene: VCA0064
A; Map position: 2

```

Query Match      7.2%; Score 290; DB 2; Length 713;
Best Local Similarity 22.2%; Pred. No. 1.7e-12;
Matches 186; Conservative 130; Mismatches 310; Indels 210; Gaps 42;

Qy 6 LKPIVLSILINTPLLAQHAEETQSVGLTETVVVKG-SRPRATSGLLHTSTASKIIIGD 64
Db 3 LSPYSAAVLISVLAAGFAHA-ETESHVEEVVVVTANRIEQ-----LSEVAGSVAVLEGE 55

```

QY 65 TL-RQKAVNLGDALDVGPGIHASQVGGGASAP---VIRGTGRIKVLNHHGTGTD---M 117

nb 56 TT-FQKGKTELVDALNCEPQGVSVT---GGARPNQITIRGTGTGNRIAVIRDGIOADGYCA 112

```
QY      118 ADFSPDHAINVDITALS-QQVEILRGFVTTLLYSNGNAGLVADVADGKIPE----- 165
ST      119 ADINDKVGNTFTSI.SNVKOTIOVVGCASSTI.YGSGAIGGVY-IIESKAPDYLTHREDYYVD 171
```

	QY	166	-XMPENGVS-----GELGLRLSSGNLEKLT------GGINIGLKG 199
	S4	170	: : : : : : : : : : : : :
		172	AATTCCTCAGTCCCGCAATACGCCGCATTCTVVGCGGTENENODI VAEPEVDGVNLCRSH 231

QY 200 NFVLTHTGLYRKS-----GDYAVPYRNLRKLPDPRRFANGQHRAVLGWRKRFYRRYTS 254

255 DRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDD 314

315 DDAHAHNGKPWIDLNRKRYELRAEWKQFPGF--EALRVHLNRNDYHDEKAGAVEN 372

[illegible]

DB	36.3	FAYGNSVDGHDYQRPKSIKRMSS-----GJLLQADEPTAPA-----KEIRFGVI	400
QY	433	GVEQANWNDETLEGGVRVEKQKASTRYDKALIDRENNYKQPLDLAGAHTQARSFALSGN	492

DB	408	QGDNLLGLDWTLAGLRFDAQKLSPKN----	TDRIGHYK--VWTWGS--SEWSPSASISYQ	466
QY	493	WYFEPQHKLSLTASHQERLPSTQELYAHGKH--	VATNTFEV-GNKHINKERSNNTLEALG	549

DB	461	WH--PEWNTVLSYNGHGRAPSYDKAYCASDHSFVPLTPFIKPNKRLAETSDFSFLGSK	518
QY	550	YEGDRQWYNLALYRNFPGNYLYAQTLLNDGRGPKSIEDDSEMKLVRY-NQSGADFYGAEGE	608

Db	519	YDNGQTQFYVAVFYSFDNFIDVKQGYDNATGSV----	IQQYQNIAGVKTYGAEMS	57
Qy	609	IYFKPTPRYIGSVGSDYVVRG-----LKNLPSLPG----	---REDAYGNRPFFTAQDD	653

Db	572	VMHRLDDRWSVENKLGVDGKDXENQVYRLLTLPLEGSVQLNVQRWDAY-----	621
QY	654	QNAQPVPAARLGLGFHLKASLTDRIDANLDYYRVFAQNKLARVETRTTQGHMNLGNVYRN	713

DB	622	-----SRLNW-----ASAMSRVPT-----CTTEQKETEBCAT-TTGWVSWDIGLNY---	661
QY	714	TRYGEWNVYVKAD-----NLLNQSVYAHSSFLSDTP-----QMGSRFTGGVNVKF	758

```

Db      :|||
662  ---QWNAQLSASFNVNLLDREYTRYQDVAGTVPSDTLYSTEPGRFT--VHAKY  711

```

RESULT 6
H90977
hypothetical protein ECS2792 [imported] - Escherichia coli (strain O157:H7, su
C:Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: H30977

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.; G Sawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. *Genome Res.* 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genome analysis

A;Reference number: A99629; UID:21156231; PMID:11258796

A;Accession: H90977

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-715 <RAY>

A;Cross-references: GB:BA000007; PIDD:BA836215.1; PID:gl3362260; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RMD 0509952

C;Genetics:

A;Gene: ECs2792

Query Match 6.6%; Score 264.5; DB 2; Length 715;
Best Local Similarity 22.5%; Pred. No. 1.1e-10;
Matches 169; Conservative 118; Mismatches 328; Indels 137

```
QY      11 LSILLINTPLLAQAHAETEQSGLETVTVVVVGKSRPRATSGLLHHTSTASDKIISGDTLRQKA    70  
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:  
  
Db     15 LSLCISGVTSATATSSETKSINSEETLVTTNR---SASNLEWSPATIOODOOTLONST    71  
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
```

Qy 71 -VNILGDALDGVPGIHASQYG-
GASAPVIRGTGERRIKVLNHHGE-TGDMADFSFDHAIM 127

b6 72 NASTANLONIDPVEITNSIAGEKQIRTEGRASSEVI.I.TDGQEVTVORAGDNNGVGUHL 131

b7c 73

QY 128 VDTALSQOVEILRGPEVTLLYSNGNVAGLYDVADGKIPEKMPENGVSQGLRLSSCNLEK 187

QY 188 LTSGGINICLGKNEVLHTEGLYKSGDYAVPRYENLKRLPDPSRRFANGQHVAVLGWRKR 247

QY 248 FYRTYSRRDOYGLPAHSHEYDCHADIIWQSLINKRYLQLYPHLLTTE-----DVDY 302

QY 303 DNPGLSCGFHDDDDAHAAHANGKPWDLRNKRYELRAEWKQPPQCFEALRVHLNRNDYHH 362

```

DD DGDYK ---KLFHDAFEIQRKQFAN---EVALIQVFSPNQLQALVH-NKLD-IR 33
290 DGDYK ---KLFHDAFEIQRKQFAN---EVALIQVFSPNQLQALVH-NKLD-IR 33
363 DEKAGDGV--ENFFNNQTONARIELRHQPIRLKGSWGVOY---LGKSSALSATSEAV 416
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

DB	338	DQYQTAVTLQSHFSLPANNELVT	----	GAQYKQDRVSRQSGMTSSKSLT	338
QY	417	KQPMLLDNKQHYSPFGEQA	----	DNFTLEGGVRVEKQKASIRYDKALID	465

DB 385 G---PINKETRTSRYSSEOSTVSLFAQNDWQFADHWTWINGVR--QYWLSSKLTGRG--D 433

QY 466 RENVYKQPLDGLAHQRTA--RSPALSGWVYFTPQKLSLTASHOER--IPSTQELVYAHG 521

QY	522	KHVATNTTFVGNKHLNKERSNNIELALGYEGDRWQYNLALYRNRFNGYIYAQTLNDORGP	581
DB	438	GVSYTAGIISDTSLARESASDHDMVTSTSLRYSGFDNLELRAAFAGQYVFPTLSQLFMQT	499

Db	498	SAGGSVTY--GNPDLKAEHNNFELGARYNGNWQLIDSAVYVSEAKDYI--ASLICDGSIV	554
QY	582	KSIEDDSEMKLVRY--NQSGADFYGAEGEYIFKPTPRYRIGVSGDIVVRGL	630

DB	555	CNCGNTSSRSYYVDNIDRAKTWGLE-----ISAEYNGWVFPISGNLIRROYETST	608
QY	631	-----EDAYGNRPFTIAQDDQNA--PRVPA-ARLGFH	667

Db	609	LKTTNTGEPAINGRIGLKHVLVMGQANIISDVFI	RAASAKDDSNCTETNVPGWATLNFA	668
QY	668	LKASLTDRIDANLDYRV-FAQNKLA--RYET	696	

Db 669 VNTEF-----GNEDQYRINLALNLTDKRYRT 695

RESULT 7
F85824

1003

Db 528 -YIDFDELQVVS-NDVGTWNLGA-----TKHQIETSGHYDFAALDPRLDGSLVYG 577
 Qy 629 RLK-----NLPSLPGREDAYGNRPFIQAQQDNAPRPAABLGHKLKASLTDRIANL 680
 Db 578 SLTYTRATYEGDIFSGRDLPLYSR-----QVATAGVRYEV-----DRWTYNL 621
 Qy 681 DYRVFAQNKLYRTETPGHMLNLGANYERN-----TRYGE-----WNVYKAD 726
 Db 622 D---AFAQSM-----QAPAGLSTDSQG-NFTHNYITEPSADGQYGDIPGYVTNARVGYD 672
 Qy 727 -----NLINQSVYAHSS 738
 Db 673 FGPOASNLKGLGVKNLFDKQYFTRSS 699

RESULT 9
 H70347
 outer membrane protein c - Aquifex aeolicus
 C;Species: Aquifex aeolicus
 C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
 C;Accession: H70347
 R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
 Nature 392, 353-358, 1998
 A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A;Reference number: A70300; MUID:98196666; PMID:9537320
 A;Accession: H70347
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-680 <AQF>
 A;Cross-references: GB:AE000694; NID:g2983162; PIDN:AA06764.1; PID:g2983163; GB:AE000694
 A;Experimental source: strain VFS
 C;Genetics:
 A;Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
 F;45-219/Domain: tonB-dependent receptor amino-terminal homology <TNN>

Query Match 6.0%; Score 243; DB 2; Length 680;
 Best Local Similarity 21.0%; Pred. No. 3.5e-09;
 Matches 165; Conservative 117; Mismatches 326; Indels 178; Gaps 33;

Qy 29 QSVGLETVVVGKSRPRATSGLLHTSTASDKIISGDTLRQ-KAVNLGDALD-GVPGIHAS 86
 Db 17 QENVLEEIQVIGKR-----EVLTEDEVRELPAKDVGAEINIPGVWKV 60

Qy 87 QYGGASAPVIRGOTGRRIKVNLNHGETGDMADFPDHAIM-VDPLASQQVEILRGPVTL 145
 Db 61 RKGAIANDVIRGFKKDEVNQLFDGARVYNACPNRMDPGIEHVPDFSEVESVEVIKGPFD- 119

Qy 146 LYSSGNVAGLVADGKIPKMPGVSGELGRLSSGNLEKLTSGGINIGLGNFVLHT 205
 Db 120 VRYGAVGTGVN-----KTEPKGMEGRITTYADNWS-----TINPSYFSGKDRUSFL 171

Qy 206 EG-LYKSGDYAVPRYNLRLPDSPPRPAHQ-----HRAVLGWRKRYRRTYS 254
 Db 172 IGVAFFRGKPYEDGKGRKITEIYPSGNPSAYSDQDERDSTAFNIHTA---WAKLAYKIDG 228

Qy 255 DR-RQYGLPAPSHSEVDDCHADILINQKSLINKRYLYQLPHLLTE---EDVDYDNPGLSCG 310
 Db 229 VRKFLDY---AHQRATD-----VLYPYLMDGIYDEVDRVNLGLB-- 265

Qy 311 FHDDDDAHAAHNGKFWIDLRNKRKYELRAEWKQPPFGFEALRYHLNRNDYHDEKAGD-- 368
 Db 266 -GKRPEAKLYGSSVRHWTNPKRVISQNAAPRGVTGTYAKSVYGFKEGY---SFGDFS 320

Qy 369 -AVENFNQQTQARIELRHQIFIGRLKSGWGVQYLGQKSSALTSATSEAVKQPMLLDNKQV 427
 Db 321 FGIDTFYRYWKAQTMYMENMGMYR-----TQNTIPDVVDVY 356

Qy 428 HYSFFGVEQANWD-NFTLGGVYRVEKQKASIRYDKALIDRENYKQPLPLDGAHQRTARS 486
 Db 357 NFGLYGEYRKLSPKLRILVAGLRDLWSKTKADSGKANILNRYH-----HNTDTTS 407

RESULT 10

H83055
 probable outer membrane hemin receptor PA4710 [imported] - Pseudomonas aeruginosa (str.
 C;Species: Pseudomonas aeruginosa
 C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 23-Dec-2002
 C;Accession: H83055
 R;Stover, C.K.; Pham, X.Q.; Brwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Li
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat
 A;Reference number: A82950; MUID:20437337; PMID:10984043
 A;Accession: H83055
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-764 <STO>
 A;Cross-references: GB:AE004885; GB:AE004091; NID:g9950968; PIDN:AAG08096.1; GSPDB:GNO
 A;Experimental source: strain PA01
 C;Genetics:
 A;Gene: PA4710
 C;Superfamily: iron-regulated outer membrane protein

Query Match 5.7%; Score 231.5; DB 2; Length 764;
 Best Local Similarity 19.8%; Pred. No. 2.8e-08;
 Matches 172; Conservative 126; Mismatches 325; Indels 247; Gaps 35;

Qy 6 LKPIVLSILIN-----TPLLAQAHETQSVGLETVVVGKSRPRATSGLLHT 53
 Db 9 LRP-CLALLLSPLSALAGNAVPLTPTTITATTEQAVDSVPSTVSVQTRQLD----- 61

Qy 54 STASDKIISGDTLRQKAVNLGDALDGVPGIHASQVGGASAPV---IRGQTRRIKVLN 109
 Db 62 -----RQVNNIKELRYEPGV--SVGGORAGITGYNIKIDGNRLTQI 106

Qy 110 HHGETGD--MADFPDHAIMVDPLASQQVEILRGPVTLTSSGNVAGLVN----- 157
 Db 107 DGVLPNDPFGSDPYAQTHRYNVDPIVKRVEILRGPSALYGSNAIGGASVYFTLDPSDI 166

Qy 158 VADGKIPKMPGVSGELGRLSSG---NLEKLT-----GGINIGLGNFV 202
 Db 167 IKDGG-----DVGARLKAGYESASHLSWLTATVAGRADDFDGLLHYGRQGHE 214

Qy 203 LHTEGLYRKSGDYAVPRYNLRLPDSPPRPAHQHRAVLGWR-----KAFYRR 251
 Db 215 TESNGGCGTG-----LSKSEANPEDADSYLLGLGNVYAGSRFGLVFEKYS 265

Qy 252 TYSDRRDQYGLPAHSHEYDDCHADI-----IWQK---SLINKRYLYQLPHLLTEE 298

```
Db      266 VDTDKSAYGGP-----YDKGKPAIPSPMLPGGMYQMRKGNLTLTREYRGLHFFILDSQ 320
QY      299 DVVDNPGSLSCGHDDDDAHAAHAKGKPMIDLNRKYELRAEWKQPPFGFEALRVHLNRN 358
Db      321 VADRIQWLSNLYQLAKTDQATREFY-----YPIRKKVLRT 355
QY      359 DYVHDEKA---GDAVENFNNTQNA---RIELRHQPIGRLKSGWQVQLGQSSALSA 411
Db      356 DTTIKERLWVFDQDLKSPAIGTEHLLSYGINKHOKVTKMRSGTGNLDTGADSPRDA 415
QY      412 TSAVKQPMLLDNKVQHYSPFGVEQANWDFLLEGGVRVE-----KQKA 455
Db      416 LERSSDFP---DPTVKYALFAQDSISWNTWTFPGLRYDYTRMEPHITDEFLTKQSQ 472
QY      456 STRYKALIDRENYKQPLDLAGHRTARSFALSGNWYTPQHKLSLTASHOERLPSTQ 515
Db      473 NTAVDDES---DKMKHRVSPKFG---VTYDFAQHYTWY---GQYAOGF-----RTPAK 516
QY      516 ELYAHGKHVATNTFEVGNKHLNKNRSNIELALGYEGDRWQYNALYRNFNGYIYAQTL 575
Db      517 ALYGRFENLOAGHYIENPNLKEKQSFETGLRGKFDGSGFVAVFYNYRDFIDEDAL 576
QY      576 N-----DGRGPKSIEDSEMKLRYNOSGADF-----YCGARGEIYFKPTPRYRIGVSG 623
Db      577 NTDSTGCGQTFCQSNRIERAVIK---GVELKGRLELGAFGAQGLYTQGSVAYAYGRNK 632
QY      624 DYVRGR-LKULPSLP-----GREDAVGNRPFIAQDDONAPRVPAARLGHFKASLT---D 674
Db      633 D--NGEPPINSVNLTPGVFGYDEADGN-----YGGLLSWTLVKRKD 672
QY      675 RIDANLDYRVFAQNKLYRVTTPGHHMLNGLANYR-----RNRTRYGEWN 720
Db      673 RVDDSS-----TFHTPDGTASQFTPGFGLVLDLSAYRSLKDLTLNAGLYNLTKDKYWLMD 727
QY      721 WYVKADNLLNQSVYAHSSFLSDTPQMGRSF 750
Db      728 DVRGYDSVGEASALAPAN-IDRLSQGRNF 756

RESULT 11
Db2317
iron-regulated outer membrane virulence protein, TonB receptor family VC0475 [imported]
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82317
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
charlson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: D82317
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-652 <HEI>
A:Cross-references: GB:AE004134; GB:AE003852; NID:g9654900; PIDN:AAF93648.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0475
A:Map position: 1
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog

Query Match 5.7%; Score 231; DB 2; Length 652;
Best Local Similarity 21.6%; Pred. No. 2.4e-08;
Matches 176; Conservative 108; Mismatches 310; Indels 222; Gaps 40;

QY      1 MAQTLLKPIVLST---LLNTPLLAQ-AHETEQSVGLTWTVVVVKGRPRATSGLLHTSTA 56
Db      1 MSRENPSPVSLVTLGLMFSASAFADATKDTETM---VVTAGVYAO-----VIONAFA 51
QY      57 SKKISGDTLRQKAV-NLGDALDGVPGIHASQVGGASAPV-IRG-----Q 100
```

```
Db      52 SISVISREDLESRYRDVTDALKSVPGVTVT---GGGTTTDISIRMGSENVTLLILDGKRQ 109
QY      101 TGRRIKVLNHHGTGDMADFPDPAHIMVDFTALSQOVELRGPVTLTLLSSGNVAGLVADVAD 160
Db      110 TSQTRP-NSDGPQIEQWMLPLQAI-----ERIEVIRGPMSTLYGSDAIGGVINIIT 161
QY      161 GKTPKMPENGVSQELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGVYAPRY 220
Db      162 RKDQOQW-----SGNVQLSTVVOEN-----RASGD----- 186
QY      221 RNLKRLPDSRRFANGQHRVAVLGRKRFYRTYSDRRDOYGLPAHSHEYDDCHADIIWQK 280
Db      187 -----EQSANFFVTGSLDALS--LQVYQQTQORDEDEI-----EHGYGDKSL-----R 228
QY      281 SLINKRYLQLYP--HLLTEEDVD---YDNPGLS-----CGFHDDDDAAHAH---AH 322
Db      229 SLTSKLYQLNPDHQLQLEAGVSAQDRENNVGKSAQSSGCGRTCSNTDQNYRRNHVAVSH 288
QY      323 NGKPEWIDLNRKYELRAEWKQPPFGFEALRVHLNRNDYH--HDEKAGDAVENFNNQTON 380
Db      289 QG-----DWQD-----VQSDTYLYQYBENTNKSREMSIDNTVFK 322
QY      381 ARTIELRHQPIGRUKSGWQVQLGQSSALSATSEAVKQPMLLDNKVQHYGFFGVQANWD 440
Db      323 STLV---APIGEHLGFGVE--GKHESLEDKTSNKISSRTHISN-TQWAGFIEDWALAE 376
QY      441 NFTLEGGVRVEKQASIRYDKALIDRENYKQPLDLAGHRTARSFALSGNWYTPQHK 500
Db      377 QFRLTFGGRLDHDK-----NY-----GSH-----FSPRVYGVNLDPLWT 411
QY      501 LSLTASHQERLPSTQELYAHGKHVATNTFEVGNKHLNKNRSNIELALGYE-GDRWQYNL 559
Db      412 VKGGVSTGRAPOLRRVTPDWGQVGGGNIYGNPDLKPKETSIKLSLMTSTGSLAASL 471
QY      560 ALYRNRFP-----GNVYIAQTLNDGRGPKSIEDSEMKLVRYNQSGADFVGARGEI 609
Db      472 TAFHNDFKDKITFVACPANICTAGPNQWGAAP-----TYRVNIDBAETYGABATL 521
QY      610 YFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPFFIAQDDONAPRVPAARLGFHLK 669
Db      522 SLPITESVELSSSYTYTHSEQKS-----GNFAGRPLL-----QLPKHLFNANLS 565
QY      670 ASLTDRID--ANLDY--YRVFAQNKLYRVTTPGHHMLNGLANYRNTRYGEMWNVYKA 725
Db      566 WQTTDLRLNWANLNYRGKMQEGGASNDDFIAPSYTFIDTGVYALTQ-----TATIKA 620
QY      726 D--NLNLQSV-YAHSSFLSDTPQMGRSFTGGVNVKF 758
Db      621 AVYNLFQEVNVAEYGVVED---GRRYWLGLDIAP 652

RESULT 12
AD0160
probable outer membrane protein YPO1313 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AD0160
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0160
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-665 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90143.1; PID:g15979363; GSPDB:GN00175
C:Genetics:
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homo

Query Match 5.6%; Score 228; DB 2; Length 665;
```

Best Local Similarity 20.7%; Pred. No. 4e-08;
Matches 152; Conservative 116; Mismatches 280; Indels 188; Gaps 34;
QY 24 AHETQSYGLFETVTVVGRSPRATSGL---LHTSTASDKIIISGTLQKAV-NIGDLDG 79
Db 29 AEKNTATPTDMV-----TASGFQORIQDASAISVVVTRQIENKAYRDVTDALD 81
QY 80 VPGIHASQYGGASAPV-IRGQTRIKVLAHGHGTGMADFP-----DHAIMVDTA 131
Db 82 VPGVVIT---GGSTSDISIRGMAKYTLIL-VNGKRVTRTRPNSDGSQTEQWLPPLA 138
QY 132 LSQQVEILRGFVTLVSYSGNAGLVADVADGKIPEK-----MPENGVSGBGLRLS 181
Db 139 AIDRIEVVGRGPMSSLYGSDAMGCVINIIRKVGKHWGTVRADATLQBDSSGDI----- 193
QY 182 SGNLEKLTSGGINIGLKNFVLTHTGLY-RKSGDYAVPRYRNKLPDPSRPFANGQRA 240
Db 194 -FQTNAYASGLTDLGL---LGLVSGLLSHRSEDKIIDGY-----NQORMRNGTATP 241
QY 241 VLGWKRFYRTYSRDRDQYGLPAHSHEYDCHADIIWQSLINKRYLQLYPHLLTEEDV 300
Db 242 TL-----TPDDNNEFPDIGHVQDRNSTPGRTLALNG---TNSDT 279
QY 301 DYDNFGLSCGTHDDDAHNAHNGKFWIDLRNRYELRAEWKQPPGFALRVHLNNDY 360
Db 280 QYDR-----NNYAVTHNG---YYDFGNS-----TSYIQRDET 308
QY 361 HHDERAGDAVENFFNQTNARIELRHOPIGRLKSGWGVYLGQSSALSATSSEAVKQPM 420
Db 309 RNPQRMSVDNLFN--TQTSFLDNHTLI--LGGQYRYEELYDKGNQLPSASDL----- 359
QY 421 LLDNKVQHYSPFGVEQANW---DNFTLEGVRYVEKQKASIRYDKALIDRENYKQPLPDL 477
Db 360 ---KKLTRMSWALFAEDQMTNDFTALTGIRMQD-----QNY----- 395
QY 478 GAHROTARGFALSGNWFYFPOKHLISLTASHQRLPSTQ-----ELYAHGKHVATNTE 530
Db 396 GTH-WTPR---LYGVWHLADQWLTKGVSQGYRSPDLRQATDDWQLSGGKG-CLPALI 450
QY 531 VGNKHLNKRNSNIELALGYG-DRWQYNLALYRNFNYI-----YAQTLD 577
Db 451 LGNSLNKPSRSIQEIGILMDQEGNMASVTLYTDFDKDITEVNCDDITNTTQCCVFN 510
QY 578 GRGPKSIEDSEMKLVYRNSQADFYGARGEIFYKPTPRYRIGVSGDYVGRGLKNLPSLP 637
Db 511 GINYKFISS-----RINVDKAMTGAETAFANDINQAWSLATWYTFQSEKSS----- 558
QY 638 GREDAYGNRPFFIAQDDQNAAPRVPARLGFHLKASLTDRIDANLDYRVFPAQNKLAYPE-- 695
Db 559 ---GAFAGQL-----NQMPKHLNGTLNKTTEDEFATWIRAN---YRGKASBYLNRTSMG 608
QY 696 TRTPGHHMLNIGANYR 711
Db 609 SRTPSYTFVDLGNAYQ 624

RESULT 13
B64049
outer membrane protein hxC homolog - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: B64049
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Pine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: B64049
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-744 <TIGR>

A;Cross-references: GB:U32696; GB:L42023; MID:g1573057; PIDN:AAC21789.1; PID:g1573065;
Query Match 5.6%; Score 228; DB 2: Length 744;
Best Local Similarity 19.9%; Pred. No. 4.8e-08;
Matches 173; Conservative 125; Mismatches 318; Indels 252; Gaps 42;
QY 7 KPIVLISILLINTPLLAQAHEHQSGVLETVTVVVKSRPRATSGLLHTSTASDKIIISGDTL 66
Db 7 KRIFLVLTVGLQVNTAKQNSNSNRELLPI-----IVNTDSDNKLPGRSVL 56
QY 67 RQK-----AVNLGDALDGVPGIHASQYGGASAPVIRGOTGRIRKIVLAHGHGTGMADFP 120
Db 57 KQNIHQXQADNAANILNLPVGNMA---GGFRP---GGGT-----LNING-MGDAEDV 103
QY 121 S-----PBAIMVDTALSQQVEILRGFVTLVSYSGNAGLVADVADGKIPEKM 167
Db 104 RVQLDGATKSPFKYQOGSIFIEPELLRRVTVDKGNVSPQYNGGFGAGTVEKFEKTDARDEL 163
QY 168 PEN-GVSGEIGLRLSGNLEKLTSGGINI-GLGNFVLHTGLYKRSQDYAVPRYRNKLR 225
Db 164 QENQKIGFLKYGNNSNNQKTYSTALVLQNEQKNIDLLFGSVRNAGDY-----KR 215
QY 226 LPDSRPFANGQRAVL-----GWRKRFYRTYSRDRDQYGL 262
Db 216 PDNSKILFSGNQKQTKGLIKLNQISPEHLLTSSVYGIHKGW-----BPPAKRDLIPK 269
QY 263 PAHSH--EYDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSGCFHDDDDAHAAH 320
Db 270 PSLSDIMRYG---TDIAWKKLVYR-----DQKQENYTL-----KYN 304
QY 321 AHNGKPIDLRNRYELRAEWKQPPGFALRVHLNRNDVHHDEKAGDAVENFFN----- 375
Db 305 LPENNPNINLSTQ-----FYSKT---TQNDMPKPEASSGLVSGLGNOSWIT 348
QY 376 -----NOTONARIELR-HQPIGRKLSGWGVYLGQSSALSATSSEAVK----- 417
Db 349 YSLLTPDINNTSTFNKTTVHELL-----FGLQWLKNTNTLMYDKSKVRKADYNYGF 402
QY 418 QPMLLDNKVQH-YSPFQVEQANWNTFLEGV-----VEKQKASIRYDKALIDRENY 470
Db 403 QPYTMPSGROYTQAFYLDQDIKWNIIFSTGVRYDHINIGQKNLAKYN----- 452
QY 471 KQPLDPLGAHQVARSFALSGNMY---FTPOHKLISLTASHQE---RLPSTQELYA---HG 521
Db 453 -----DISAGHDYSQKNYNGWSYLLGLYDNDVHLSLFTNFSKTWRAPVIDEQYEQFKQ 507
QY 522 KHVAINTFFGVNKHNLKERSNNIELALGY---EGDRWQYNLALYRNFPGNYI----- 571
Db 508 SSVFATSLNLEKEMINOTRYGGI-ITLNLHFQENDAFQFETTYFYNRGKNEIFKTRGVNC 566
QY 572 ---AQTLDNGRGPKEISDDSEMKLVRYNQSGADFYGARGEIFYKPTPRYRIG-VSGDYVR 627
Db 567 VGNADTNNKVCPLIEN-----YRNLPGYVIOGAELEYQST---YLFGELTYSYVK 617
QY 628 GRLKNLPSLP-GRBDAYGNRPFFIAQDDQNAAPRVPARLGFHLKASLTDRIDANLDYRVF 686
Db 618 GKRDTSRPNWGTSTW-----IA---EIPPRKATTAALGENVPKY-----YLTVG 659
QY 687 AQNKLAVERTRP-----GHEMLNIGANYRNTRYGEW-----NMYVKADNLNQ 731
Db 660 WRABFVRQRDRSPSLSGDPKASSWSLPSRGSYLSLNLFLSWSPAKIKGMNVKITVDNLNFR 719
QY 732 SVYAHSSFLSDTPQMGSRFTG-GVNVKF 758
Db 720 AY-----NPVIGELASGTGRNIKF 738
RESULT 14
S25265
N;Alternate names: ferrienterochelin receptor homolog
C;Species: Vibrio cholerae
C;Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 08-Oct-1999

C;Accession: S25265; A37834
R;Goldberg, M.B.; Boyko, S.A.; Butters, J.R.; Stoeber, J.A.; Payne, S.M.; Calderwood, Mol. Microbiol. 6, 2407-2418, 1992
A;Title: Characterization of a Vibrio cholerae virulence factor homologous to the family A;Reference number: S25265; MUID:93023868; PMID:1406279
A;Accession: S25265
A;Molecule type: DNA
A;Residues: 1-652 <GOL>
A;Cross-references: GB:U72152; EMBL:M63192; NID:g1763224; PIDN:AAC44766.1; PID:g1763226
A;Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 299-1Thz
R;Goldberg, M.B.; Boyko, S.A.; Calderwood, S.B.
J. Bacteriol. 172, 6863-6870, 1990
A;Title: Transcriptional regulation by iron of a Vibrio cholerae virulence gene and homolog
A;Reference number: A37834; MUID:91072235; PMID:2174861
A;Accession: A37834
A;Molecule type: DNA
A;Residues: 1-152, 'D', <GO2>
A;Cross-references: GB:M37773
C;Genetics:
A;Gene: irgA
C;Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo
C;Keywords: membrane protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-652/Product: outer membrane protein irgA #status predicted <MAT>
F;68-214/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F;367-652/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

	Query Match	5.6%; Score 227; DB 2; Length 652;
	Best Local Similarity	21.6%; Pred. No. 4.6e-08;
	Matches 176; Conservative 108; Mismatches 310; Indels 222; Gaps 40;	
Qy	1 MAQTILKPIVLSI---LLINTPLLAQ-AHTEGVSGLTEVTTVVGKSRPRATSGLLHTSTA	56
Db	1 MSRFNPSPVSLVTGLMFMSASAFADATKTDETM---VVTAAGYAO-----VIQNAPA	51
Qy	57 SDKIISGDTLRKAV-NLGDALGVPGIHAQSYGGGASAPV-IRG-----Q	100
Db	52 SISVLSREDELEGRYRDVTLAKSVPGVTVT---GGDTTDSIRKMGSNYTLILVDGKRQ	109
Qy	101 TGRRIKVLNHHGETGDMADFPSPHAINVDTALSOQVEILRGPTVLLTSSGNVAGLVAD	160
Db	110 TSRQTRP-NSDGPFIGQGLPPLQAI-----ERIEVIRGPMSTLYGSDAIGVINIIT	161
Qy	161 GKIPKMPGVSGELGLRLSSGNLKLKLTGGINIGLKNFVLHTEGLYKSGDYAVPRY	220
Db	162 RKDOQQW-----SGNVQLSTVVQEN-----RASGD-----186	
Qy	221 ENLKELPSPRPFANGQRAVLGWRKRFYRTVSDRDQYGLPAHSHEYDDCHADIWQK	280
Db	187 -----EOSANFVTGTPLSDAL- -LQVYGQITQRDEI-----EHGYGDKSL-----R	228
Qy	281 SLINKRYLQLYP- -HLLTSEEDVD- -YDNPGLS-----CGPHDDDDAHAAH- -AH	322
Db	229 SLTSKLNQLNPDHQLQLEAGVSAQDRENNVGHSAQSSGCGRTCSNTDNQYRRNHVAVSH	288
Qy	323 NGKPWIDLNRKRYELRAEWKQFPFGPEALRVHLNRNDYH- -HDEKAGDAVENFNNQTON	380
Db	289 QG-----DWQ-----GVCQSDTYLQYBENTKSRMSIDNTVFX	322
Qy	381 ARIELRHPIGRKLSGWQVYLQKSSALSATSEAVKQPMLLDNKVQHYHFFGVEQAWND	440
Db	323 STL- -APIGHEMLSFGE- -GKHESLEDKTSNKISSRTHISN- -TQWAGFIEDENALAE	376
Qy	441 NFTLEGGVRVEKQKASIRYDKALIDRENYVQPLPDLAGHRQTARSFALSGNWVFTPQHK	500
Db	377 QFRLTFGGRLDHDK-----NY-----GSH-----FSPRVTVGMNLDPLNT	411
Qy	501 LSLTASHQERLPSTQELYAHGKHVATNTPEVGNKHLNKRSNLIALGYE- -GDRWQYNL	559
Db	412 VKGGVSTGTFRAPQLREVTPDMQVSGGGNIYGNPDLQPETSINKLSLMTYSTGSLAASL	471
Qy	560 ALYRNRF-----GNYYIAQTLDNRGPKSIEDDSMKLVRYNQSGADFYGAEI	609


```

Db      106 SQILYHQR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAI 146
Qy      165 EKMP-----NGVSGELGLRLSG--NLEKLTSGGINIGLGNFVLTGELY-----RKSQ 213
Db      147 AKTVDAQDILKGLDKWGVRLNSGFASNEGVSYGASVFGKEGFP-----DGLFSYRNDEK 202
Qy      214 DYAVPR-YRNL---KRLPDS-----RRFANGQHRVAVLGRKRFYR-----251
Db      203 DYAGKGFNVNGKTPVPSALDKRSYLAKITTFGDDHRLVLSHMKQHRGIRTVRE 262
Qy      252 -TVSDRRDQYGLPAHSHVEDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNFGLSCG 310
Db      263 FTVGDKSSRINIDROAPAYR-----TTQSTNLTAYTGKILG 299
Qy      311 FHODDDAHAHNGKFWIDLRNKRY---ELRAEWKQFPFGFEALRVHLNRNDYHDEKAG 367
Db      300 FVEKLDANAYV-----LEKERYADSGTGAGYAGNVKGPNEHRTITTCGANFNFDRLA 351
Qy      368 DAVENFNQOT-QNARIELRHQ---PIGRUKSGWGVQYLGQKSSALSATSEAVKQPM---420
Db      352 E-----QTLKYGINVRHOBKIPQAFNLKSFSIPTTEKN-----GQKVDKPMEOQ 397
Qy      421 -----LLDNKQCHYSPFFGVEQA-NWDNFTLEGGVREVEKOKASIRYDKALIDR 466
Db      398 MKDRADEDTVHAYKLSNPTKIDTGVVVEALHDIGFTLTGLRYDRFKVK--THDGKTVSS 456
Qy      467 ENYKQPLDGLAHROTARSFALSGNNYFTPOHKLSLTASHOERLPSTOELVYAHGKH---523
Db      457 SNLN-----PSFGVIWQPH-----HWSFSAHNY---ASRSPRL--YDALQTHGKRGII 501
Qy      524 -VATNTEVGNKHLNKNERSNNIELALGYEDGRWQYNLALYRNFNGYIYAOTLNDGRGPK 582
Db      502 SIADGT-----KAERAETGIFNYNDGTFAAN-----GSYFW-QTKDALANP 544
Qy      583 SIEDDSMKLVRYNOSGADPYGAE-GEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRED 641
Db      545 QNRHDSVAVREAVNAGYIKNHGYELGASVYRTGGLTAKGVVS-----HSKPRFYD 593
Qy      642 AYGNRPPIAQQDQVAPRPAARLGFHLKASLTDRI--DANLDY-YRVFAQNKLYARTTRTP 699
Db      594 THKDKLLSANPEF-----GAQVGRTWASLAYRFQNPFLNLEIGWR-----GRYVQKAT 640
Qy      700 GHMLNLGANVRR-----NTRYGEW-----NVVYKADNLNLSVVAHSS 738
Db      641 G-SILAGQDKRKNLENVVRKGFVNDVFNANKFKLQKDTLNLSNVNPNFKFYTPHSQ 699
Qy      739 FLSDT-FQMGSRFTGGVNVKPF 758
Db      700 RWTNTPGVGRDVRGLGVNYPF 720

RESULT 18
F81962
probable iron-regulated outer membrane protein NMA0453 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: F81962
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: F81962
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-714 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CA883750.1; PID:g737920
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: fetA; NMA0453
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
Query Match 5.3%; Score 213.5; DB 2; Length 714;

```

```

Best Local Similarity 22.0%; Pred. No. 4.9e-07;
Matches 188; Conservative 103; Mismatches 308; Indels 257; Gaps 47;

Qy      11 LSILLINTPL-LAAHETEQSGLETTVTVCKSRPRATSGLLTSTASDKIISGDTLRQK 69
Db      8 LSLISLTLAAGFAHAENNANVALDVTIVKGDQ-----GSKRTNIV--TLQOK 55
Qy      70 ----AVNLGDALDGVPGIHASQYGGGASAPVI-----RGQTGRRIKV-----LNHH 111
Db      56 DESTATDMRELLKEEPSI---DFGGNGTQSFLTRGMGQNSVDIKVDNAVSDSQILVHQ 112
Qy      112 GETGMADFPDPAIMVDTALSQQVEILRGVPTLLYSNGVAGLVVDVADGKIPKMPB---169
Db      113 GR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIIAKTVDQA 153
Qy      170 ---NGVSGELGLRLSG--NLEKLTSGGINIGLGNFVLTGELYRKS-----GDYAVPR- 219
Db      154 DLLKGLDKWGVRLNSGFASNEGVSYGASVFGKEGFP-----DGLFSYRNDEKYEAGK 209
Qy      220 YRNL-----KRLPDS-----RRFANGQHRVAVLGRKRFYRRTYSRRDQY--GLP 263
Db      210 FRNPNFGKTPVPSALDKRSYLAKITGTFGDDHRLV-----SHMKQHRGIR 258
Qy      264 AHSHEYDDCHADIIWQKSLIN-KRYLQLYPHLLTEEDVDYDNFGLSCGFHDDDAHAHAH 322
Db      259 TVREEF-----TVGGDKERISMKQAPSYR--TTQSTNLTAYTGKDLGFVEKLDANAYV- 311
Qy      323 NGKPWIDLNRKY---ELRAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFNQOT- 378
Db      312 -----LEKKRYSDSGSGVAGNVKGNPNTQITTCGANFNFDRLAE-----QTL 356
Qy      379 QNARIELRHQ---PIGRUKSGWGVQYLGQKSSAL-----SATSEAVKQPMLLDNKVQHY 429
Db      357 LKYGINYRHOBIKQAFNLNSQFKIE--DKEKATDEKKKNRENEKIAKAYELTNPTKD 413
Qy      430 SPFGVEQAN-WDNFTLEGGVREVEKOKASIRYDKALIDRENYVKQPLDGLAHROTARSFA 488
Db      414 AGAYIEALHEIDGFTLTGGLRYDRFKVTHDGK-----TVSSSS 452
Qy      489 LSGNW--YFTPOHKLSLTASHO--ERLPSTOE-LYAHGKH-----VATNTEVGNKHLNKE 539
Db      453 LNPSFGVIWQPHHEHWSFSAHNSYASRSLYDALQTHGKRGIIISADGT-----KAE 504
Qy      540 RSNNTIELALGYEDGRWQYNLALYRNFNGYIYAOTLNDG-RGPKSIEDSEMK-----591
Db      505 RARNTGIFNYNDGTFAAN-----GSYFW-QTKDALANPQNRHDSVAVREAVNAGY 555
Qy      592 -----LVRYNOSGADPYGAEGEIYFKPTPRYRIGVSGDYVRG---R 629
Db      556 IKNHGYELGASVYRTGGLTAKGVSHSKPRFYDTHDKLLSANPEFGAQVGRTWASLAYR 615
Qy      630 LKNLPSLP---GR---EDAYGNRPPIAQQDQVAPRPAARLGFHLKASLTDRIDANLDY 683
Db      616 FXN-FNLEIGWGRYVQKAVGSILVAGQKRSGLKLENVVRQGVN-----660
Qy      684 RYFAQNKLYARTTRTPGHHMLNLGANVRRNTRYGEMWYKADNLNLSVVAHSSFLSD 743
Db      661 DYFANKPKLGKDTLN-----VNLSVN-----NVNFTFYTPHSQRWNT 698
Qy      744 -FQMGSRFTGGVNVKPF 758
Db      699 LFGVGRDVRGLGVNYPF 714

RESULT 19
D81215
TonB-dependent receptor, probable NMB0293 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: D81215
R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.
Hickey, S.K.; Haft, D.H.; Salzberg, S.D.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

```

[illegible][illegible]

```

Db 500 GT-----KAERARNTIEGNYNDGTPAAN-----GSYFW-QTIKDALANPQRKH 542
Qy 586 DDESEMK-----LVRNOSGADFYCAEGEELYFKPTPRYRI 619
Db 543 DSVAVREAVNAGYIKNHGYELGASVRTGGLTAKVGVSHSKRPFYDTHDKLLSANPEFCA 602
Qy 620 GVSQGYVVG-----RLKNLPSLP-----GR--EDAYGNRPPIAQDDQONAPRPAARLGFHLKA 670
Db 603 QVGRWTASLARYFQON--PNLEIGWGRYVQKAVGSILVAGQKDRNGKLENVVRKGFQVNV-- 660
Qy 671 SLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGLGANYRNTRYGWNWVYKADNLLN 730
Db 661 -----DVFAWKKPLGKDTLN-----VNLSVN-----NVFN 685
Qy 731 QSVYAHSSFLSDT-PQMGSRFTGGVNVKF 758
Db 686 TFYPYHSQRWINTLPGVGRDVELGVNYKF 714

RESULT 21
A56148
pepticin receptor [imported] - Yersinia pestis
C:Species: Yersinia pestis
C>Date: 03-Oct-1995 #sequence revision 03-Oct-1995 #text_change 11-May-2000
C:Accession: A56148; T17444; T47052
R:Fecherston, J.D.; Lillard Jr., J.W.; Perry, R.D.
J. Bacteriol. 177, 1824-1833, 1995
A:Title: Analysis of the pepticin receptor from Yersinia pestis: role in iron-deficient
A:Reference number: A56148; PMID:95204350; PMID:7896707
A:Accession: A56148
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-673 <FE2>
A:Cross-references: GB:U09530; NID:G493075; PIDN:AAA69906.1; PID:G493076
R:Fecherston, J.D.; Bertolino, V.J.; Perry, R.D.
Mol. Microbiol. 32, 289-299, 1999
A:Title: YbTP and YbTQ: two ABC transporters required for iron uptake in Yersinia pestis
A:Reference number: Z18782; PMID:99248409; PMID:10231486
A:Accession: T17444
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-673 <FE2>
A:Cross-references: EMBL:AF091251; NID:G3818595; PID:G3818609; PIDN:AAC69592.1
R:Buchrieser, C.; Rusniok, C.; Couve, E.; Frangeul, L.; Billault, A.; Kunst, F.; Carniel
submitted to the EMBL Data Library, October 1998
A:Description: DNA sequence of the 102 kbases unstable region of Yersinia pestis.
A:Reference number: Z24348
A:Accession: T47052
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-673 <FE2>
A:Cross-references: EMBL:AL031866; PIDN:CAA21395.1
A:Experimental source: strain 6/69
C:Genetics:
A:Gene: psn
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
C:Keywords: membrane protein
F:66-202/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:382-672/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 5.2%; Score 211; DB 2; Length 673;
Best Local Similarity 21.8%; Pred. No. 6.7e-07;
Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

Qy 1 MAQTLKPIVLSILINPLLAQAHETEQSVGLETVTVVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTRLYPLALGGLL--PAIATAQTSQD---ESTLVVYASKQSSRSA--SANNVSTV 53
Qy 61 ISGDTLRQKAVNLGDALGV--PGIHASQYGGGAGAPV--IRGTGRRIRKIVLNHHGETGMA 118
Db 54 VSAPELSDAGVTASDKLPRVLPLGLNIENSGNMLFSTISLRGVSSAQ----- 99

```

```

Qy 119 DF-SPDHAIMVD-----TALS--QQVEILRGPTVTLVSSGNVAGLVADVADKIPB 165
Db 100 DFYNDAVTLVYDGVLPQLSTNTIQTALTDVQSVELLRGPQGTLYGKAQGGIINVTQQ--PD 158
Qy 166 KMP-----ENGVSGLGLRLSSNLKLTSGGINIGLGNFVLHTEGLYK--SGDYAVPR 219
Db 159 STPRGYIEGGVSSRSYR--SKFNL-----SGPIQDGL-----LYGSVTLRQVDDGDMINFA 209
Qy 220 YRN-----LKRLLPD-----SPRRFANGQHRVILGWRKFRVRRYTSR 256
Db 210 TGSDDLGGRASIGNVKRLAPDDQPMWGPFAASRECTATQDAYGVN----- 258
Qy 257 RDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEBDVDYDNPGLSCGFHDDDD 316
Db 259 -----DIKGRKLSISDGSDDPYMRCT-----DSQTLGKYVTTDD-- 293
Qy 317 AHAHAHNGKPIDLRNKRYELPAEWKQ-----PPGFALRVLHLNRNDVHDEK----- 365
Db 294 -----WV-----FNLISAWQOQHYSRTFFS--GSLIVNMPQR--WNQDVQELRAAT 335
Qy 366 AGDA--VENFFNNQTONARIELRHQPIGRLGKSGWGVQLGQKSSALSATSEAVKQPMLLD 423
Db 336 LGDARTVDMVFLYQNTN-----EKLNSAYDMPYPLSSYGYTTAEFLAA----- 382
Qy 424 NKVQHYSPFGEQANW-----DNFTLEGVVRVEKQKASIRYDKALIDRENYKQPLDPLGAH 480
Db 383 -----YS-----DLTWHLTDRFDIGGVRFSHDKSSTQVHGSMLG-----NPFQDQGS 426
Qy 481 RQTARSAFALSGNWYTPPHQKLSLTASHQERLPSTQELVAHGKHVATNTPFVGNKHLNKR 540
Db 427 NDDQVLGQLSAGYMLTDDMRV--YTRVAQGYKPSGYNI-----VPTAGLDAKPFVAEK 477
Qy 541 SNNIELALGYEGDRQVYNLALYRNRGNVYIAQTLDNGRGPKSIEDDSEMKLVRYNOSGA 600
Db 478 SINYELGTRYE-----TADVTIQAATF--YTHTKDMLVSGPVRMOTLS-----NAGKA 524
Qy 601 DFYGAAGIYFKPTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNRPPIAQDDQONAPRV 660
Db 525 DATGVELEAKRFRAPGMSWDINGNVIRSEFTN-----DSELYHGKR-----VFPVP 570
Qy 661 AARLGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGLGANYRNTRYGE-- 718
Db 571 RYGAAGSSVNGVIDTRYGA-----LMPRLAVNLVGPYFP--GNQLRQGTATLD 618
Qy 719 -----W-----NWYVKADNLLNQ--SVYAHSSFLSDTQMGSRFTGGVNVK 757
Db 619 SSLGWQATERMNIIVYVNDLFDRRYRTGYWNGSSAVAQVNMGRVGVGINTR 669

RESULT 22
AF0232
pepticin/yersiniabactin receptor protein [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AF0232
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.I.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
ii, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413; PMID:11586360
A:Accession: AF0232
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-673 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90722.1; PID:G15979925; GSPDB:GN00175
C:Genetics:
A:Gene: fyvA
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
Query Match 5.2%; Score 211; DB 2; Length 673;
Best Local Similarity 21.8%; Pred. No. 6.7e-07;
Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

```

Mol. Microbiol. 13, 253-263, 1994

A>Title: The pesticin receptor of *Yersinia enterocolitica*: a novel virulence factor w

A|Reference number: S49044; MUID:95075311; PMID:7984105

A|Accession: S49044

A|Molecule type: DNA

A|Residues: 1-44 <RAW>

A|Cross-references: EMBL:Z29675

C|Genetics:

C|Gene: fyuA

C|Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homol-

C|Keywords: membrane protein

F|1-22/Domain: signal sequence #status predicted <SIG>

F|23-673/Product: iron-repressible outer membrane protein fyuA #status predicted <M

F|66-202/Domain: tonB-dependent receptor amino-terminal homology <TN>

F|382-672/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 5.2%; Score 210; DB 2; Length 673;

Best Local Similarity 21.7%; Pred. No. 7.9e-07;

Matches 180; Conservative 95; Mismatches 320; Indels 236; Gaps 43;

Qy 1 MAQTTLKPVLVSILLINTPLLAQAHETQSGVLEIVTVVVKSRPRATSGLLHTSTASDKI 60

Db 1 MKMTRLPLALGGLL--PAIATAQTSQDSELTFTV--ASKQSSRSAS----ANNVSSTV 53

Qy 61 ISGDTLRKAVNLGSDALDGV--PGTHASQYGGGASAPV--IRGQTGRRIKVLNHHGETGDMA 118

Db 54 VSAPELSDAGVTASDKLPRLVPLGLNIENSGNMLFSTISLRGVSSAQ----- 99

Qy 119 DF--SPDHAIMVD-----TALS--QQVEILRGPTVLLYSSGNVAGLVADVADGKIPE 165

Db 100 DFYNPAVTLYYDGVQPLSTNTQIALTDVQSVELLRGPGQTLYGKSAQGGIINIVTQQ--PD 158

Qy 166 KMP---ENGVSSELGLRLSSNLEKLTSGGGINIGLGNFVLHTEGLYRK--SGDYAVPR 219

Db 159 STPRGYIEGGVSSRDSYR--SKFNL-----SGPLQDGL---LYGSVTLLRQVDDGDMNPA 209

Qy 220 YRN-----LKRLPD-----SPRRFANGQRAVLGWKRKFRYRRYSR 256

Db 210 TGSDDLGGTRASIGNVKRLRAPDDQPEWMGFAASRECTRATQDAYVGMN----- 258

Qy 257 RDQYGLPAHSHEYDCHADIWQKSLINKRYLYQLPHLLTEEDVDVDPCLSGCFHDDDD 316

Db 259 -----DVKRKLSDGSPDPYMRCT-----DSQTLGSKYTTDD- 293

Qy 317 AHAHAHGKPIDLRNKRKYELRAEWKQ-----PFPGFALRVHLNENDYHDEK----- 365

Db 294 -----WV-----FNLSAQQQHYSRTFFS--GSLIVNMPQR--WNOQVQELRAAT 335

Qy 366 AGDA--VENFPNQTQNARIELRHQPIGRLKSGWGVQYLGKSSALSATSEAVKQPMILD 423

Db 336 LGDARTVDMVFLGYEQNTR-----EKLNSAYNNPTPYLSSTGYTTAETLAA----- 382

Qy 424 NKVQHSYFPGVEQANW---DNFTLEGGVVERVEQKASIRYDKALIDRENYKQPLDLGAH 480

Db 393 -----YS-----DLTWHLTDRPDIGGVRFVSHDKSSTQYHGSMLG-----NPFQDGKS 426

Qy 481 RQTARSPALSGNNYPTPOKUSLTSASHQBRLSPTQBYLAHGKHVAANTTEVGNKHLNKR 540

Db 427 NDDQVLGLQSAGYMLTDWVRV--YTRIAQYKPKSGYNI-----VPTAGLDKAPFAEK 477

Qy 541 SNNIELALGYEDRWQYNLALYRNFQYIYAQTLLNDGFGPKSIEDDSEMKLYRNOGSA 600

Db 478 SINYELGTRYE-----TADVTLOAAATF--YTHTKDMQLYSGPGVMQMTLS-----NAGKA 524

Qy 601 DIFYGAEGEYFKFPTPRYRIGVSGDYVRGRKLNPLSLGREDAYGNRPFFIAQDDQNAFVRP 660

Db 525 DATGVELEAKWRFAPGWSNDINGVIRSEFTN-----DSELYHGNR-----VPFVFP 570

Qy 661 AARLGFHLKASLTDRIDANLDYRYFPAQNKRLARYETRTPGHFMNLGANYRVRNTRIGE-- 718

Db 571 RYGAGSSVNGVITDTRYA-----LMPRLAVNLVGPYFD--GDNQLRGQGTATLD 618

Qy 719 -----W-----NWYVKADNLLNO--SVYAHSSFLSDTPQWGRSFPTGGVNVK 757

Db 619 SSLGWQATERINISVHVDLFDRLRYRTYGYMNGSSAVAQVNMGRVTGINTR 669

RESULT 24

E81792

Probable TonB-dependent receptor protein NMA2193 [imported] - Neisseria meningitidis (st C) species: Neisseria meningitidis

C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C/Accession: E81792

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, L.; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000

A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A/Reference number: AB1775; MUID:20222556; PMID:10761919

A/Accession: E81792

A>Status: preliminary

A/Molecule type: DNA

A/Residues: 1-703 <PAR>

A/Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85405.1; PID:g738081

A/Experimental source: serogroup A, strain Z2491

C/Genetics:

A/Gene: NMA2193

Query Match	5.2%	Score 208.5;	DB 2;	Length 703;
Best Local Similarity	20.6%;	Pred. No. 1.1e-06;		
Matches 181;	Conservative 119;	Mismatches 259;	Indels 321;	Gaps 52;
Qy 12 SILLINTPLLA---QAHETEQSGVLEVTVVVKSRPRATSGLLH-----TSTASDKIISG- 63	Db 12 TLIIASFVAADTQNGEHYATLPVSVVVGQDTSVLKGYINVDEAAVTRNGQLIKET 71			
Qy 64 ----DTLR-QKAVNLG-----DALDGVPGIHASQYGGGASAPVIRGQTRRIKVLNHHGE 113	Db 72 PQTIDTLNIQKNKYGTNDLSSILEGNAGIDA-----AYDMRGES-----IFLRGF 117			
Qy 114 TCDMADFSPDHA-----TWVDTALSQQVEILRGVPTLLYSSGNVAGLVADVADGKIPKVP 168	Db 118 QADASDIYRDGVRESQVRRSTANIERVELKPSGLYVTRNGGGVINMVSKVAFKQS 177			
Qy 169 EN-----GVSGELGLRLSS--GNLEKLTSG--GINIGLGNFVLHT 205	Db 178 RNIGTVYGSWNRSLMNDINEVLKNVAIRLTGVEGRANSFRSGIDSKNMVWVSPSITVKL 237			
Qy 206 EGLYRKSGDYAVPRYRNLRKLDP--SPRRFANGQRAVLGWKRFRVRRYTSDDRQOYGLP- 263	Db 238 DNLGKWGTQYT--YDNVERTPDRSP-----TKSVY-----DRFGLPY 272			
Qy 264 ----AHSHEYDCHADIIWOKSLINKRYLOLYPHLLTEEDVDYDNPGLSCGFHDDDDAHA 319	Db 273 RMGFARNDP-----VKDKLQW-----RSDLEY----- 296			
Qy 320 HAHNCKPWIDLRNRYELRAEWKOPFPFGPEALRVHLNRNDYH--DEKAGDAVENFF-- 374	Db 297 -AFNDK-W-----RAQWQ-----LAHRTAAQDFHFYAGSENGLIKRNVAW 336			
Qy 375 ---NNQTONQVARIELR-HQPIGRLKG--SWGVOYLQKQ-----SSALSAT-- 412	Db 337 QQTDNKTLSSNLTLDNGDYTIGRPFENHLTVGMDYSREHRNPTLGFSSAFASINPYDRASW 396			
Qy 413 -SEAVKQPMLLDN--KVQHYSPF--GVEQANWD-NFTLEGVVRVEKQASIRYDKALIDR 466	Db 397 PASGRLOPILTONRHKADSYGIFVQNIFFSATPDLKFLVGG-----RYDKYTFNS 445			
Qy 467 ENYKQPLDLAGHRO-TARSPA--LSGNWYFYPQHKLSLTASHQERLPSTQELIYAHGKH 523	Db 446 ENKL-----TGSSRYGSHSPFNIGAVWNINPVH--TLIYASYNK--GFAPYGGRGY 494			
Qy 524 VATNTFVGNKELNKNERSNNIELAL--GYEGDRWQYNLALYRNFRFGNVIYAQTLDNGRGP 581	Db 495 LSIDTLSSAVFNADPEYTRQYETGVTKSSWMLDRLSTLSAY----- 535			
Qy 592 KSIEDDSEMLVRYNQSGADFGAEGEYFKPTPR---YRIGVSGD----- 624				

```
QY 445 EGVVVEKQASIRYDKALIDRENYKQPLDPAHROTAARSFALSGNNWYFPPQHLSLT 504
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 QAGVHSHMKMEV-----DDRYLSNGDASGRYRKNTSPSPVMYAFPTDLHGYS 470
QY 505 ASHQBRLPSTOEL-YAHGKHVATNFEVGNKHLNKNERSNNIELALGYEGDRW--QYNLA 560
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
471 AKGGETTQEMAYA---PVANAPDVFNFGKLPATSSQYB--AGLKARLMGNTRVNA 525
QY 561 LYRNRFNGYIYAQTLDNGRGPKSIEDDSEMKLVRYNQSGADF-----YCAEGEI 609
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
526 IFQVTEDEIVVASSLGR-----TSQNAGKTLRRGFELGLESELSSEHWN 572
QY 610 ---YFKPTPRIGV-SDGYVRGRLKNLPSLP-----GREDAYGNRP 647
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
573 NLAYTRLSATVSDSDEAGGKTIGKGLHFGVPESLFGELVWKPAGISMGWEGMYRSQV 632
QY 648 FIAODDQAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKARAYETRTPGHMLNLG 707
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
633 YV--EDNSEKAAPS-----YAVF--NWRTRFE----- 656
QY 708 ANYRNTRYGENWY--VKADNLNQSVYAHSSFLSDTPQM-----GRSFTGGVNVKF 758
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
657 -----QRLGAWTFHQLVRLDNLPRQ-YVGSVIVDGNRRYEAAPGLSMVAGAGVEY 708

RESULT 26
T12069
hemin receptor - Yersinia pestis
C:Species: Yersinia pestis
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T12069
R:Horning, J.M.; Jones, H.A.; Bertolino, V.J.; Perry, R.D.
Submitted to the EMBL Data Library, October 1998
A:Description: hmukSTUV of Yersinia pestis encodes an ABC transport system necessary for
A:Reference number: Z17401
A:Accession: T12069
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-676 <HOR>
A:Cross-references: EMBL:U06447; NID:g3776552; PID:g1407797
A:Experimental source: strain KIM6
C:Genetics:
A:Gene: hmur
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
C:Keywords: membrane protein

Query Match 5.1%; Score 205.5; DB 2; Length 676;
Best Local Similarity 19.3%; Pred. No. 1.7e-06;
Matches 164; Conservative 110; Mismatches 282; Indels 293; Gaps 37;

QY 11 LSILLINT-PLLAQAHETEQSVGLETVVVGKSRPRATSGLLHTSTASDK----- 59
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 LSLAIACLTPLATQA-----ADTTTQTSSKKHSTDVTMTATGNERSSEFAPMMVT 65
QY 60 IISGDT-LRQKAVNLGDALDGVPGIHASQYGGASAPVIRGQTGRRIKVLNHHGDTGM- 117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 VIEGNAPTSQTAATAADMLRQVPLTVTG-----SRTNGQDVVMRGYKQGVLT 115
QY 118 -----ADSPDHAIWDTALSQQVEILRPVTLTLYSSGNVAGL-----VDVADGKIP 164
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 LVDGVRQGTDTGHLNSTELDPALVKRIEIVRPAALLYGSGALGVIAETVDAAD---- 171
QY 165 EKMPEGVSGELGRLSSGNLEKLTSGGINTGIGKNFVLHTE-----GLYRK 211
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 --MLQPGQNS--GYRVYS-----SAATGDHSPFLGASAFGRDSDLGLSLFGTRDIGNIRQ 223
QY 212 SGDVAVPR-----YRNKLELPSPRRFANGQHEAVLGW 244
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224 SNGFNAPNDEITSNVLAKGTWQIDSIQSLSANLYRNNSALEPKNPQTSAPSSTN-VMTN 282
QY 245 RKRFRVRRYSRRQYGLPAHSHEVDDCHADIIWKSILNKRYLQLYPHLLTEEDVDYDN 304
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
283 RSTIQR-----DAQLRYNIKPLDQELNATAQVYSEVEINAR-----PQGSAREGREQTT 333
```

```
QY 305 PGLSCGFHDDDAHAHANGKPIWIDLRNKRVELRAEWKQPPGFEALRVHL---NRNDYH 361
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
334 BGVK-----LENRTFLF-----IESPASHLLTYGTETYK 362
QY 362 HDEKAGDAVENP-----FNNQTONARIELRHQPIGRILKSGWGVQYLGOKSALSATSEA 415
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
363 QEOTFGGATESFPQAKIRFSSGLWQDEITLRLDPLVSIILAGT----- 403
QY 416 VKQPMLLDNKVQHVSFFGVGEQANWDFNILEGGVRVEKQKASIRYDKALIDRENYKQPLP 475
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
404 -----RYDNYSGSSDGY-----A 416
QY 476 DLGAHROTARSAFALSGNNWYFPPQHLSLTASHQE--RLPSTOELVAGHKA-- 525
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
417 DVDADKWSRG-AIS---ITFTDMLFLGSAQAAPRAFTWCMYVNDSKHFAIPRPLGT 471
QY 526 -TNTEVGNKHLNKNERSNNIELALGYE-----GDRWQYNLALYRNFGNYIYAQTLD 577
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
472 LTN-YWVPNPNLKPETNETQYGFGLRPSDLLMAEDDLQFKVSYFDTRAKADYISTRV--- 527
QY 578 GRGPKSIEDDSEMKLVRYNQSGADFYGAEGEYFKPTPRYLIGVSGDVYRGRLKNLPSLP 637
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
528 -----DMQAMTTTSSVNIQAKIWDGMSYK-TALFNWDLAYNTRGNQN----- 573
QY 638 GREDAYGNRPFTIAODDQANPR-----VPAARLGFHLKASLTDRIDANLDYRVFAQNK 691
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
574 --TDEM-----LDTINPDVTISVDVPSVANSFP-----SVGMGTGTA-NRS 611
QY 692 ARYETRTTP--GHHMLNLANGYRNTRYGENWYVKADNLNQSVYAHSSFLSDTPQMGRS 749
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
612 SRVSSSTPQAGYGVNDFYVSYKGQAFKGMITMLLGNVFEKEYTPOGI-----PDGR- 666
QY 750 FTGGVNVKF 758
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
667 -----NVKRF 670

RESULT 27
AG0035
hemin receptor precursor [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AG0035
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.F.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0035
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-676 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89146.1; PID:g15978384; GSPDB:GN00175
C:Genetics:
A:Gene: hmur
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog

Query Match 5.1%; Score 205.5; DB 2; Length 676;
Best Local Similarity 19.3%; Pred. No. 1.7e-06;
Matches 164; Conservative 110; Mismatches 282; Indels 293; Gaps 37;

QY 11 LSILLINT-PLLAQAHETEQSVGLETVVVGKSRPRATSGLLHTSTASDK----- 59
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 LSLAIACLTPLATQA-----ADTTTQTSSKKHSTDVTMTATGNERSSEFAPMMVT 65
QY 60 IISGDT-LRQKAVNLGDALDGVPGIHASQYGGASAPVIRGQTGRRIKVLNHHGDTGM- 117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 VIEGNAPTSQTAATAADMLRQVPLTVTG-----SRTNGQDVVMRGYKQGVLT 115
QY 118 -----ADSPDHAIWDTALSQQVEILRPVTLTLYSSGNVAGL-----VDVADGKIP 164
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 LVDGVRQGTDTGHLNSTELDPALVKRIEIVRPAALLYGSGALGVIAETVDAAD---- 171
QY 165 EKMPEGVSGELGRLSSGNLEKLTSGGINTGIGKNFVLHTE-----GLYRK 211
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 --MLQPGQNS--GYRVYS-----SAATGDHSPFLGASAFGRDSDLGLSLFGTRDIGNIRQ 223
QY 212 SGDVAVPR-----YRNKLELPSPRRFANGQHEAVLGW 244
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224 SNGFNAPNDEITSNVLAKGTWQIDSIQSLSANLYRNNSALEPKNPQTSAPSSTN-VMTN 282
QY 245 RKRFRVRRYSRRQYGLPAHSHEVDDCHADIIWKSILNKRYLQLYPHLLTEEDVDYDN 304
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
283 RSTIQR-----DAQLRYNIKPLDQELNATAQVYSEVEINAR-----PQGSAREGREQTT 333
```



```

Db 116 LVDGVRQGTDTGHLNLTFLDPAVKRIEIVRGPAALLYGSGALGGVIAETVDAAD----- 171
QY 165 EKPENGVGELGLRLSSGNLEKLTGGGINIGLKNFVLHTE-----GLYRK 211
Db 172 --MLPFGQNS--GYRVYS-----SAATGDHSFGLGASAFGRDLDLGLSFGTRDIGNIRQ 223
QY 212 SGDYAVPR-----YRNKLKLPDPSRRFANGOHRAVLGW 244
Db 224 SNGFNAPNDETISNVLAGTWTQWIDTSQSLSANRURYNNNSAIEPNPOTSAPSSTN-VMTN 282
QY 245 RKFYRTYSRRDQYGLPAHSHEDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDN 304
Db 283 RSTIQI-----DAQLRYNRPDQEWLNATAQVYSEVINAR-----PGSAEGRQIT 333
QY 305 PGLSCGCFHDDDAHAHANGKPMIDLRNKRKYELRAEWKQFPFGFHALRVHL-----NRNDYH 361
Db 334 EGVK-----LENTRLF-----IESPASHLLTYGTETYSK 362
QY 362 HOEKAGDAVENF-----FNNQTONARIELRHOPIGRLKSGWGVQYLGQKSSALSATSEA 415
Db 363 QEQTPGATESPQAKIRFSSGWLQDEITLRLPVSILAGT----- 403
QY 416 VKQPMLLDNKVQHSYSPFGVEQANWMDNFTLEGGVRVEKQKASIRYDKALIDRENIYKQLP 475
Db 404 -----RYDNVSGSSDGY-----A 416
QY 476 DLGAHQRTARSFALSGWNFTPOHKLSTASHQE--RLPSTQELYAHGKHVA----- 525
Db 417 DVADAKWSRG-AIS-----ITPTDMLFGSYAQAFRAFTMGEMVNDYSKHFALPIRPLGT 471
QY 526 -TNTPFVGNKHLNKNKERSNNIELALGYE-----GDRWQVNLALYRNRFNYIAOTLND 577
Db 472 LTN-YWVPNPNLKPETNETQYGFGLRFSDDLMAEDDLQFKYSYFTKAKYISTRV----- 527
QY 578 GRGPKSIEDDSEMKLVRYNQSGADFYGAEGEIYFKPTPRYRIGVSGDYVRGRKLNPLSLP 637
Db 528 -----DMQAMTTTSVNIIDQAKIWGDASMSYK-TALFNWDLAYNRTRGKNQ----- 573
QY 638 GREDAVGNRPFFIAQDDQNAPR-----VPAARLGFHLKASLTDRIDANLDYRVFAQNK 691
Db 574 --TDEW-----LDTINPTVTSIVDVPVANSFG-----SVGWIGTFA-NRS 611
QY 692 ARYETRTP--GHHMLNLGANYRRNTRYGBWNWYVXADNLLNSVYAHSSFLSDTPQMGRS 749
Db 612 SRVSSSTPOAGYGVNDFYYSYKQAEAFKGMTTLLGNVFEKYEYTPQGI-----PODGR- 666
QY 750 FTGGVNVKF 758
Db 667 -----NVKF 670

RESULT 28
A57148
outer membrane protein hmxC - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
C>Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 22-Oct-1999
C:Accession: A57148
J:Cope, L.D.; Yoge, R.; Muller-Eberhard, U.; Hansen, E.J.
R:Bacteriol. 177, 2644-2653, 1995
A:Title: A gene cluster involved in the utilization of both free heme and hemoexin
A:Reference number: A57148; PMID:95270579; PMID:7751272
A:Accession: A57148
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <COP>
A:Cross-references: GB:U09840; NID:G595325; PIDN:AAA87059.1; PID:G595326
A:Experimental source: type b
C:Genetics:
A:Gene: hmxC
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homology
C:Keywords: membrane protein
F:59-188/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:408-725/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

```

```

Query Match      5.0%; Score 203; DB 2; Length 725;
Best Local Similarity 19.8%; Pred. No. 2.8e-06;
Matches 158; Conservative 134; Mismatches 300; Indels 206; Gaps 43;

```

```

QY 11 LSTILLINTPLAQAHTEQSVGLETTVVVCKSPRATSGLLHTSTASDKIISGDTLRQKA 70
Db 6 LSLAIAT--LVTANALQSVELDSINVIAITRDP--SRPAYTPEKQSK--DSLISKQA 57
QY 71 VNLGDALDGVPGIHASQYGGG--ASAPVIRGTGRR-IKVLNHHGTGMADFSPPH-- 124
Db 58 TSVAALALEIPVDIR--GGSRSIAQKPNIRGLSDNRVQVIVDGRQNFDA-----HRG 110
QY 125 AIMVDTALSOQVEILRGVPTLYSSGNVAGLVDAVDGKIPEKMPGVSGELGRSSG- 183
Db 111 SYPLPMSLIQEIVKIPGSSSLWGSGALGGVAMRTNPALDLDLKN--DKFGVKIRQGY 167
QY 184 -NLEKLTSGGINI-GLGKNFVLHTEGLYKRSQGYAVPRYNLKLKLPDPSRRFANGOHRAV 241
Db 168 QTANNLSESDVSFAANDKFPVLISFYNNADNLRGTGKNLKN--TAYKQFGG--LAK 222
QY 242 LGWRKFRYRT-YSDRRDQVGLPAHSHEDDCHADIIWQKSLINKRYLQLYPHLLTEEDV 300
Db 223 FGWQINDANRVELSHRETRFKOTAPN-----NEVENELTNEQIITDQIR 266
QY 301 DYNDPGLSCGFHDDDAHAHANGKPMIDLRNKRKYELRAEWKQFPFGFALRYHLNRNDY 360
Db 267 EFHKP-----NNGSP-----PKAKPSQEEF-----YSGVKTRFGSVSY 299
QY 361 HHDEKAGD-AVENFF--NNQTONARIELRHOPIGRLKSGWGVQYLGQKSSALSATSEA 415
Db 300 LTDQQLPDQSTVFNYLTPTDNPVNLTHIALYNNKT-----IEKEQRKVSQVKDQTKL 351
QY 416 VKQPMLLDN--KVQHSYF-FGVEQANWMDNFTLEGGV-----RVEKQKA----- 455
Db 352 TTRGINLRSSSELSHLSFYVGYDYM-RDKIRTERGTNNKDAQPRADPYNANSNTTGYLI 410
QY 456 -----STRYD-----KALIDRENIYKQPLDGAHQRTARSFALSGNWYF 495
Db 411 AHIPLFGEXLLSPSVRYDHYDTSSKTVKYKDNHL-----SPATKLTIWV 455
QY 496 TPQHKLSTASHQE--RLPSTQELYAHGKHVAT-----NTFFVGNKHLNKNERNNIE 545
Db 456 T--NMLDFTAKYNEAFRAFSQMOERFVSGSHFGTSLGRNEINKF-VANPNLRDETAKNKE 512
QY 546 L-----ALGYEGDRWQVNLALYRNRFNYIAOTLNDGRGPKSIEDDSEMKLVRYNQ 598
Db 513 ITANLHFDLSLFKQGDKFKIEATYFRNDVKDFNLKIFNDAKTNTNASAG-----GA 564
QY 599 GADFYGAEGEIFYKPT-PRYR-----IGVSGDYVRGRL--KNLPSLPGREDAYGN 645
Db 565 GA--GANPNGALLPTKSOYQNTNARLSGIELQAOYQTERLTFTFTNYGSTKCKDKDSG- 620
QY 646 RPFIAQDDQNAAPRPAARLGFHLKASLTD--RIDANLDYRVFAQNKLAARYTRT-PGH 701
Db 621 -----EALSNIAASKIGVGNVALVKDKFTVGATVTTHYA--AQRVPVKDHSVTYPSY 670
QY 702 HMLNLGANYRRNTRYGEW 719
Db 671 ILTDLRAITYA--PLKGEW 686

RESULT 29
AC2075
outer membrane heme receptor alr2153 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2075
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguc.
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A

```


A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: Ac2075
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW73852.1; PID:gl
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2153

Query Match	5.0%;	Score 203;	DB 2;	Length 802;
Best Local Similarity	21.0%;	Fred. No. 3.3e-06;		
Matches 177;	Conservative 90;	Mismatches 333;	Indels 242;	Gaps 38;
Qy	26	ETEOSVGLTETVTVGKSRPRATSGLLHTSTASDKIISGDTTLRKAVNLGDALDVGPGIHA	85	
Db	94	BENEADIE-ITVGTTRPSLDESFGITVID--AEDIDNQLIQLNDDLIRYEPGVST	149	
Qy	86	S-----QYGGGASAPVIRQOTGRRIKVLNHHGETGMADFDPH--AIMVDTALSQOUEIL	139	
Db	150	SGDARY--GFQDFNIRGIDGNRVLLQVDGVRLPDSFGSTGLGRNVYIDTETLRRVEII	207	
Qy	140	RGPTVLLSYSSGNVAGLVADVADGKIPEKMPENGVSCELGLRLSSGNLEKLTSGGINIGLCK	199	
Db	208	RGASATLVGSDAIGGVVTFITKPSDYLNESDDGYFS-----NKFVYDSANRGIGN	259	
Qy	200	NFVL-----HTEGL--YKSGDYAVPRYRNKLXLPDSPRFANGOHRAVLGRKE--	242	
Db	260	TTTIAGRLGDVEGLLIYTRRDGYE-PQI-NSDRAENPQTIDAN-----SWLTKLVFNL	310	
Qy	248	-----FYRRT-----YSDRRDQVGLPAHSHEDDCHADIWQKSLINKRYLQLY	291	
Db	311	GFDDQLKLTGEFINRTTDTNVLTSRGINFGVRTDSLTAID-----EIKENRYNLSY	361	
Qy	292	PHLLTEEDVDYDNPGLSCGFHDDDDAHAAHNGKPIWIDLRNKY-----ELRAE	340	
Db	362	EH-----NNPOSNLFFQ-----VLRSQIYYQEATSTSESNELERRA	396	
Qy	341	WKQPPGPFPEARVHLNRNDYHDEKAGD-AVENFENNQTQARI---ELRHOPIGRLKG	395	
Db	397	-TAPITTCGAVNRFRFSIIYQNTGGDIQLESNFQTGDINHRLVYGAELSNKTSRLRD	455	
Qy	336	SW-----GVQYLQKKSALSATSEAVKQPMLLDNKRVQHSYFFGV---EQANWDNFLE	445	
Db	456	GFOENIAAPGI--IGERTNRVGPDPAPVKDIADTDN-----TRFGVYLQNEITWGNLTII	508	
Qy	446	GGVRYEKQKASTRYDKALIDRNVYKQPLPOLGAHQRTABRSALSGNVYFPQHKLSTA	505	
Db	509	PG-----IRYD-----SYSLNFPD--DIYRANGRNPTNFSDSALSPLGLVY	551	
Qy	506	SHQERLPSTQELIYAHGKHVATNT-----FEVGNKHLNERSNNIELALYEGD	553	
Db	552	SFTSEVTGFAQ-YSRGFRAPTGEDINPGFTNPQGVTVIPNPLDKAETSNNPGLGRGSPF	610	
Qy	554	RHOYNLALYRPNFNVIA-----QTLNDGRGPKSIEDSEMKLVRYNQS	598	
Db	611	SGKFSISGYNYNPFIDTGFVRVITPGLIVSGFQTVNRG-----EVRI-----	654	
Qy	599	GADFYCAEGEYFKFTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFTIAQDDQNAP-	657	
Db	655	---YQVEAKGELKLGSGFSLLASSSYTVG-----DDLEENQPLVSDVDFRLVT	699	
Qy	658	---RYPAARLGFHLKASL-----TDRIDANLDIYRVPQAKNLARYETTPGHMLN	705	
Db	700	GURYAPQNRWGTELIGSYAAASPRLPKTEIENPFVDPDSYFTLDLIGYY-----N	749	
Qy	706	LGANYRRNTRYGEWYVYKADNLNOSVY-----AHSFSFLSDTPQMGSRFTGGVNV	756	
Db	750	LSDNATLN-----IGVFENLLNEKYWRGVDGRGLSANDGNLDLIFVQPGISLVAGLTV	800	
Qy	757	KF 758		
Db	801	RF 802		

RESULT 30

A56268

Fe-regulated protein B precursor - Neisseria gonorrhoeae (strain FA19)

N:Alternate names: outer membrane protein FnpB

C:Species: Neisseria gonorrhoeae

C>Date: 13-Sep-1995 #sequence_revision 13-Sep-1995 #text_change 22-Oct-1999

C:Accession: A56268

R:Beucher, M.; Sparling, P.F.

J. Bacteriol. 177, 2041-2049, 1995

A:Title: Cloning, sequencing, and characterization of the gene encoding FnpB, a major

A:Reference number: A56268; PMID: 7721696

B:Accession: A56268

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-713 <BEU>

A:Cross-references: GB:U13980; NID:g833694; PIDN:AAC43332.1; PID:g833695

C:Genetics:

A:Gene: fnpB

C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homol.

F:1-22/DNA: signal sequence #status predicted <SIG>

F:62-191/DNA: tonB-dependent receptor amino-terminal homology <INN>

Query Match 5.0%; Score 202; DB 2; Length 713;
Best Local Similarity 21.6%; Pred.No. 3.2e-06;
Matches 187; Conservative 97; Mismatches 306; Indels 274; Gaps 46;

QY 11 LSILLINPL-LAQAHETEQSGLTETVTVGKSRPRATSGLLHTSPASDKIIISGDTLRQX 69
||| : : : : :
DB 8 LSLLSLTLAAGFAHAENANVALDTVTIVKGDRQ-----GSKIRTNIV--TLQQK 55
||| : : : : :

QY 70 ---AVNLGDALDVGPGIHASYGGASAPVI-----RCQTGRIRKV-----LNHH 111
||| : : : : :
DB 56 DESTADIRELLKEEPSI---DFGGNGTSFQLTRGMGNQSVDIKVDNAYSQILIYHQ 112
||| : : : : :

QY 112 GETGDMADFSDPHAIWDTALTASQQVEILRGPTVLTYSSGNVAGLVADVADCKIPERKMP-- 169
||| : : : : :
DB 113 GR-----FIVDPALVKVVSQKG-----AGSASAGIGATNGAIIAKTVDAQ 153
||| : : : : :

QY 170 ---NGVSGELGLRLSSG-NLEKLSTGGINIGLKGNFVLHTEGLY----RKSGDYAVPR- 219
||| : : : : :
DB 154 DLLKGLDNKWGVRLNSFGANNASGYASVFQKGFNF---DLGSYNRNDEKDYEAGK 209
||| : : : : :

QY 220 YRN---LRELDPSP-----RRFANGOHRAVLGWKRKYR--RTYSRRRDQYGLP 263
||| : : : : :
DB 210 FENDNGKKTVPYSALDKRSYLAKIGTTFGDGHRIVLSHWKDQHGRITRV--REEPAVS 266
||| : : : : :

QY 264 AHSHEYDDCHADI IWQKSILNKRYLQLYPHLLTEEDVDYDNPCSLSCGFHDDDAHAHAHN 323
||| : : : : :
DB 267 EKN-----SRITIQRAPSRE-TTQSNLTAYTKDLGFEVKLDANAVV-- 310
||| : : : : :

QY 324 GKPWIDLNRKRYELRAEWKPQPPGFALRVHLNRNDYHHDKAGDAVENFNNOTONARI 383
||| : : : : :
DB 311 -----LEKKRYS--ADDKD-----NGYAGNVKGPNETRIATRSMNPFDS 348
||| : : : : :

QY 384 ELRHOPIGRLKSGWGVOYLGO-KSSAL-----SATSEAVKQPMLLDNKVCHYS 430
||| : : : : :
DB 349 RLAEQTL-LK--YGINTYHOIKQAFLNSPEIKDKKATNEEKQNENEKIACKAYR 404
||| : : : : :

QY 431 FFGVQANW-----DNTFTLEGVRVEVKQASIRYDKALI DRENYYKQPLDUGAH 480
||| : : : : :
DB 405 LTNPFTKTDGTGAYIEATHEDGFTLTGLGRYDFKVKTHDGK----- 445
||| : : : : :

QY 481 ROTASFALSNGNW--YFTPOHKLSLTASHQ--ERLPSTOB-LYAHGKH-----VATNTFEV 531
||| : : : : :
DB 446 --TVSSSSLNPSFGVITWPQREHWSFSASENAGRSRLYDALQTHGKRGIIISADGT--- 500
||| : : : : :

QY 532 GNKHLNKRSENIELALGHEGDRQWQYNLALYRNRFNYIYAQTINDG-RQPKIEDSEM 590
||| : : : : :
DB 501 -----KAERARTEIGFNNDGTFRAN-----GSY-FRQTIKDALPNQRHDSVAV 546
||| : : : : :

QY 591 K-----LVRYNOSGADFYGAEIGEIVFKTPTRVIGVSGD 624
||| : : : : :

```

418 YSS-----TNRSPYKAEIFSLFAEN-----NMEL----- 442
QY 408 ALSATSEAVKQPMLLDNKQVHYSPFGVQANW-----DNFTLBGGVVRVEKQKAS 456
Db 443 ----TDSTMLTFGI---RFDHHSIVG---DNWSPSLNLSQGLGDDFTLKQGI----- 484
QY 457 TRYDKALIDRENNYKQPLFDLGAHRQTARSEFALSGNWFYTPQHKLSTLTAHQERLPSTQJE 516
Db 485 -----ARAYKAPSLYQTNPNYIL-----YSKGGQ 508
QY 517 LYAHGKHVATNTEFVGNKHLNKNRSNNIETALGYEGDRWQYNLALYRNRFGNYIVA----- 572
Db 509 CYATGAGTGIGCYMMGNDLKAETSINKIEGLEFKRDGMLAGVTWFRNDRNKIEAGTVP 568
QY 573 -QTLNDRG-----PKSIEDDSEMKLVRYNQSGADFYGAEGEYIF-----KPT--- 614
Db 569 LQRNNGKTDYQWENVPKAVVEGLEGLT---NVPSQTVNNTNNVTYMLQSKNKTGER 625
QY 615 ----PRYRIGVSGDYVRGLKNLPSL-----PGRDAYGNRPFIADDDQNAPRVPAA 662
Db 626 LSIIPQYTLNSTLWQVRQDVSLSQSTFTWYGKQEPKKYDQGN-FVTGTDKQAVS--PVS 682
QY 663 RLGHFL-----KASLTDRIIDANLDYRVFAQNKLYARETRTPGHMMLNLA-NYRNRTP 715
Db 683 IVGLSATMDVTKNVSLTGGVD-NLFDKRLWRGNAQTVRDQTGTAYMAGAGAYTYNEPGR 741
QY 716 YGEWNWYV 723
Db 742 ----TWYM 745

RESULT 32
H81030
TonB-dependent receptor NMB1882 [imported] - Neisseria meningitidis (strain MC58 serog
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: H81030

```

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.
 ri, H.; Qin, H.; Vamathevan, J.; Gall, J.; Scariato, V.; Masignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755; PMID:10710307
 A:Accession: NB81030
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-725 <127>
 A:Cross-references: GB:AE002538; GB:AE002098; NID:g7227136; PIDN:AAF42216.1; PID:g72271
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB1882
 C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom

Query Match 5.0%; Score 200; DB 2; Length 725;
 Best Local Similarity 21.7%; Pred. No. 4.6e-06;
 Matches 178; Conservative 105; Mismatches 294; Indels 244; Gaps 43;

Qy	19	PLLAQAH-----ETQSVGLEVTVVGKSRPATSGLLHTST-----AS 57
Db	15	PVTAQADVSSDDPKPQESTELPTITADRTASSNDGYTVSGTHTPLGLPMTLREIPQS 74
Qy	58	DKIISGDTLR-QKAVNLGDALDVGPIHASYG---GGASAPVIRGQTGRRIKVLNHGGE 113
Db	75	VSVITSQMRDQNIKTLDRALLQATGTSRQIYGSDRAGYNVLFARGS-----RANYQIN 129
Qy	114	TGNWADSPPHATMVDLTALSCQVEILRGPTVLLYSSGNVAGLVDAVKIPEKMPENGVS 173
Db	130	GIPVADALADTG-NANTAAAYERVEVVRGAVGLDGTGSEPSATVNLVKKRLTRK-PLFEVR 187
Qy	174	GEIGLRLLSNLEKLTSGGGNIGLGNKFNVLHTEGLYKSGDYAVPYRNLEKRLPDSPRFF 233
Db	189	AFVGNVYHETLIDAVNGGSLV-----

```

QY 234 ANQHRAVLGWRKRYRTYSDRRDQYGLPAHSHEYD-----DCHADIWQSLINKRYL 288
D 221 GRGD-----SWR-----RRERSRDAELYGI-----LEYDIAPOTRVHAGMDYQQA----- 260
QY 289 QLYPHLLTREDVD-----YDNPGLSGCGFHDDDDAHAAHNGK-----PWIDLR-NK 333
D 261 -----KETADAPLSYAVDSQGVATAPGPKDNPATWNSRRHALLNFAGIEHRFNO 312
QY 334 RYELRAEW-----KQPPFGFALRVHLNNDYHHDEKAGDAVENFF--NNQTONARIE 384
D 313 DWKLKAEYDTRGRFQYGVAGVLSI-----DHNTAATDLIPGYHADPRTHSASVS 365
QY 385 L--RHQPIGR-----LKSGMWQVYLGQKSSALSATSEAV-----KQPMLLDNKV 426
D 366 LIGKYLRFREHDLIAGINGYKVASNKYGRSIIIPNAIPNAYFSTGYPQSPASQTI 425
QY 427 QHY-----SFFGVEQANWNTFLEGGVREKQKASIRYDKALIDRENYIKQPLDGA 479
D 426 PQYTRRQIGYLATRPRAADNLILGGRYTRVGTG-SYDSR-----TQGMTYVSA 476
QY 480 HRQTARS-----PALSQWYFTPOHKLSTASHQERLPSTQELYAHGKHVATNTEFVGNKHL 536
D 477 NRFTPYTGIVFDLTGN-----LSLYGSYSLFVPOQOKDEHGSYLKPVV----- 520
QY 537 NKERSNNIELALGYEGD-----RWQYNLALYRNFRFGNYIYAQTLLNDGRGPKSIEDDSEMKL 592
D 521 -----GNNLE--AGIKGEWLEGRMASAAVYRARKNNLATAA-----GRDP----- 559
QY 593 VRYNQSGADFY-----GAEGEIFYKPTPRYRI-----GVSGDYVR-----GRKLNLPSLPG 638
D 560 -----SGNTYRAANQAKTHGEWEIVGGRITPEWQIOAGYSQSKTRDQDSRLNPDSVPE 614
QY 639 RE-----DAYGNRPETADQONAPRPAARLGFHLKASLTDRIDANLDIYRVFAQNKLARY 694
D 615 RSFKLFYAY-----HPAPEAPS-----GWITGAGVRWQSEHTDTPALIRIPNPAKA 661
QY 695 -----ETTPGHHMLNLANYRNTRYGENWYVKAADNLLAQ 731
D 662 RAADNSRQKAVADIMARYFRNPRA-----ELSLVDNLFNK 699

RESULT 33
E83485
probable tonB-dependent receptor PA1271 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83485
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:2043737; PMID:10984043
A:Accession: E83485
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-616 <STO>
A:Cross-references: GB:AB004557; GB:AB004091; NID:99947204; PIDN:AAQ04660.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1271
C:Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to
Query Match 4.9%; Score 199; DB 2; Length 616;
Best Local Similarity 22.4%; Pred. No. 4.2e-06;
Matches 155; Conservative 88; Mismatches 249; Indels 199; Gaps 39;

QY 38 VYGRPRATSGLLTSTASPKIISGDTL-RQKAVNLGDALDGVFGIHASQYGG-GASAP 95
D 31 VVTRFTRACTAS---QSLAAVSVIDREDIERSQARSPELLRQYPGVSLANNNGGFGKNTT 87
QY 96 V-IRQGTGRIRKVLNHHGETGD-MADFPDPAHIMVDRLASQOVELRGPVTLIYSSGNA 153

```

```

D 88 LFLRGTSRDLVLIDGKIVGSASAGITAFQDLPEV-----LIERIEVVRGPRSSLYGSAIG 145
QY 154 GLVDV-----ADGKIPEKMPENG-----VSGELGLRSLSSNLEKLTSGGINIGLGNFVL 203
D 146 GVIQIFTRRGDQGAQKFFSAGYGTHTQTLGSGAGVSGAGN-----GWYSLGYSS---F 196
QY 204 HTEGLYRK---SGDYAVPR---YRNKLRLPDSPPRFANGQRAVLGWRKFRYRRTYSRRD 258
D 197 DTAGINTKRACTAGTGYEDRDGYRNLSGNLRCGYRFDNGL-----ELD 238
QY 259 QYGLPAHSHEYDDCHADIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSGCGFHDDDDAH 318
D 239 GTLLRAKSHN-----DYDQVFGNSGFANADGE 266
QY 319 AHAHNGK-----PW-IDLENKRYELRAEWKQP---PPGFEALRVHLN-RNDYHHDEKA 366
D 267 QNLVGRARFTFPDPFWLVTQAGRSEDKADAYQGRFYRFDTRRDSLSWQNDLTLEA-- 324
QY 367 GDAVENFFNNQTONARIELRHQPIGRKLSGWSGVYLGQKSSALSATSSEAVKQPMLLDNKV 426
D 325 GHVLTYLGVDKDEIISSEAFSVDSRLNKGWFAOYLGQVG-----RQDWQLSLRR 374
QY 427 QHYSFFGYE---QANWNTFLEGGVREKQKASIRYD---KALIDRENYIKQ-PLPOLGA 479
D 375 DDNQOFGVHDTGSAAW--GYALSDALRF-----TVSYGTAFKAPTFNELYDPDYGNDPLDA 428
QY 480 HRQTARSFALSQWYFTPOHKLSTASHQERLPSTQELYAHG-KHVATNTEF-----VG 532
D 429 ETSRSLEVLGSG-----THGWGHWA VNAFRTNVDDLLG 461
QY 533 N-----KHLNKRNNIELALGYEGDRQWYNL-ALY---NRNGVYIYAQTL 575
D 462 NDRPAPRPMQPNNDIEARIGVELVLSGQWGLMDWNNANATFLDPQNRSG-----V 515
QY 576 NDG-----RGPKSIEDDSEMKLVRYNQSGADFYGAEGEIFYKPTPRYRIG-----VSGD 624
D 516 NDGNELPRARRMFNLDRRFRERLS-LGASVH-AEGRRYDDPANKVGLGYATLDRSE 573
QY 625 Y-----VRGRKLNLPSLPGREDAYG-NRP 647
D 574 YRLNDEWRLQGRIANLFG-ADHETAYGYNQ 603

RESULT 34
D81976
probable ferric siderophore receptor protein NMA0575 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: D81976
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandrea
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: D81976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <PAR>
A:Cross-references: GB:ALJ62753; GB:ALJ57959; NID:g7379120; PIDN:CAB83866.1; PID:g7379:
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA0575
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hon
Query Match 4.8%; Score 192; DB 2; Length 725;
Best Local Similarity 21.2%; Pred. No. 1.7e-05;
Matches 174; Conservative 108; Mismatches 295; Indels 244; Gaps 42;

QY 19 PLLAAQH-----BTEQSGVLETVVVGKSPRATSGLLHTST-----AS 57
D 15 PYVAQADVSVDDPKPQSRSTELPTITVTRADRTASSNDGYTVSGTHTPLGLPMTLREIPQS 74

```


Db	7	TSRLSLLALAVSATLPTFAFATET-----MTVTATGNARSSFEAPMMVS-VIDTGAPE	59
Qy	58	DKIIISGDTLRKAVNLGDALGVFGCIHASQYGGGASAPV-IRGQTGRRIKVL-----NHHG	112
Db	60	N-----QTATSATDLRHVPGITLDGTGRINGQDVNMRGYDHRGLVLVDVGRQGT	110
Qy	113	ETGMADESPHAIWVDTALSCQVEIILRGPVTLIYSSGNVAGLV-DVADGKIPKPMEN	170
Db	111	DTGHL-----NGTFLDPALIKRVEIVRGPSALLYGSCALGVISYDTVDAK---DLLQE	161
Qy	171	GVSGBELGURL-SSGNLEKLTSGGINIGLKNFVLHTEGL-----YKSGDYA	216
Db	162	QOSS--GERVFTGG-----TGDHSLGLGASAFORTENLDGIVAMSRDRGDLSNGET	214
Qy	217	VPR-----YRNLEKRLPDSPPRFANGQRAVLGWRKRFY	249
Db	215	APNDESIINMLAKGTWQIDSAQSLSGLVRYNYNDAREPKNPQTVGASESNPMVDRSTIQ	274
Qy	250	RRTYSRRDQYGLPAHSEYDDCHADIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSC	309
Db	275	R-----DAQLSYKLPAQGNLDNLADAKIYWSVRINAQ-----NTGSSG	313
Qy	310	GFHDDDDAAHAHNGKPWIDLNRKRYELRAEWKQFPFGFEALRYVHLNRNDYHHDEKAGDA	369
Db	314	EYRQIYTKGAELEN-----RSTLFAD-----SFASHLLTYGGEYRQEHPPGA	357
Qy	370	VENF-----FNNQTONARIELRHQPIGRLKGSGVQYLGQKSSALSATSBAVKQPMLLD	423
Db	358	TTGFPQAKIDPSSGWLQDEITLRLDLP-----	384
Qy	424	NKVQHYSFPGVEQANWONFTLEGGRVREKQKASTRYDKALIDRENYKQPLDLAGRHQT	483
Db	385	-----TLGGTRYDSYRGS-----SDGYK-----DYDADKWS	411
Qy	484	ARSPALSGNMYFTPQHKLSLTASHOE--RLPSTOELYAHGKHVATNTFE---VGNKHLN	537
Db	412	SRA-----GMTINTNMLWLFSGYAQAFRAPTMGEMYNDSKHSIGRFTYNYWVPNLR	466
Qy	538	KERNNNIELALG-----YEGDRQWNALALYRNRFNGYIYAQTINDGRGFKSIEDDSEM	590
Db	467	PETNETQYGFGLRFDLMLNSDALEFKASYFTDKADYI-----SITVDFAA	515
Qy	591	KLVRYNOSGADFYCAEGEIVPK-PTPRYRIGVSGDYVRGLKNLPULPGREDAYGNRPFI	649
Db	516	TTMSYNVPNAKIWG--WDVMTKYTTDLFSLDVAYNRTRGK-----DSTDGEYISSINPDT	568
Qy	650	AQDDQNAAPRYAARLGHFLKASLTDRIDANLDYRVFAQNKLARYETRTFGHEMLNLGAN	709
Db	569	VTSTLN--IPIAHSGF-----SVGWVGTFADRS-----THISS	600
Qy	710	YRRNTRYGWNWYV	723
Db	601	YSKOPGYGVNDFYV	614

RESULT 37
E86022
outer membrane heme/hemoglobin receptor [imported] - Escherichia coli (strain O157:H7, s
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E86022
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotberg, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference numbers: A85480; WUID:21074935; PMID:11206551
A:Accession: E86022
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-660 <STO>
A:Cross-references: GB:AE005174; NID:g12518206; PIDN:AAGS8641.1; GSPDB:GN00145; UWGP:Z49
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:

A:Gene: chuA
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homology

	Query Match	4.7%; Score 188; DB 2; Length 660;
	Best Local Similarity	19.8%; Pred. No. 2.8e-05;
	Matches 157; Conservative	96; Mismatches 281; Indels 260; Gaps 34;

QY	4	TTLKPIVLSILLINT-PLLAAQHETEOSVLGLETVTTVVKSR-----PRATSGLIHTSTAS	57
DB	7	TSLRLSLLALAVSATLPTFAFATET-----MTVTATGNARSSFEAPMWS-VIDTSAPE	59
QY	58	DKIISGDTRLRKAVNLGDALDGVPCIHASQVCGGSAPV-TRGQTGRIKVL-----NHHG	112
DB	60	N-----QTATSADLLRHVPGITLDGTRNGQDVNRGVDHGVLVLPDVGVRQGT	110
QY	113	ETGMADPSPDHAIMVDTALSCQVEILRGPPVTLLYSNGNVAGLV--DVADGKIPEKMPE	170
DB	111	DTGHL-----NGTFDPALIKRVEIVRGPSSALLYGGGALGVISYDITDAK---DLLQE	161
QY	171	GVSSELGLRL-SSGNLEKLTSGGINIGLKNFVLHTEGL-----YRKSGDYA	216
DB	162	QQSS--GFRVFETGG-----TGDSHLGLGASAFGRGTENLDGIVAWSSRDGDLRQSNET	214
QY	217	VPR-----YRNLRKLPLDSPRRFPANGQHRVAVLGWRKEFY	249
DB	215	APNDESNMLAKGTWQIDSQAQSLGSLRVYYNNDRAREPKNPQTVCASESSNPMDRSTIQ	274
QY	250	RRTYSRRDQVGLPAHSHEYDDCHADIWKQLINKRYLYPHLLTEEDYDYNPGLSC	309
DB	275	R----DAQLSYKLAPOQNMDLNADAKIYVSEVRINAQ-----NTGSSG	313
QY	310	GFHDHDDAHAAHNKGKWIIDLNRKYELRAEWKQPFPGFEALRVHLNNDYHHDEKAGDA	369
DB	314	EYRQITKGARLEN-----RSTLFAD-----SFASHLLTYGGEYTRQSHPGGA	357
QY	370	VENF-----FNNTONARIELHQPIGRUKSGWGVOYLQGOKSALSATSBAVQPMLLD	423
DB	358	TTGFPQAKIDFSSGWLQDEITLRLPI-----	384
QY	424	NKVQHYSEFFGEQANWFNFTLEGGRVBEKKQASTRYDKALIDRENYTKQPLDIAHRQT	483
DB	385	-----TLGGTRYDSYRGS-----SDGYK---DVDADKWS	411
C	484	ARSFALSGNWYFTPQHKLSLTASHOE--RLPSTOELVAHGKHVATNTFE---VGNKHIN	537
DB	412	SRA-----GMTINPTNLMFLGSAQAFRAPRTMGEMTNDSKFPSIGRFYNYWVFNPLR	466
QY	538	KERSNNTELALG-----YEGDRMQNIALRYNRFGNIYVAQTLNDGRGPKSIDDSEM	590
DB	467	PETNETQYEGFLRGFDLLMSLNDALEFKASYFTDKAKDYI-----STTVDFAAA	515
QY	591	KLVRYNOSGADFYGAEGEIYFK-PTPYRIQVSGDYVRGRILKNLPILPGREDAYGNRPFI	649
DB	516	TTMSYVNPNAKIWG--WDVMWKYTDLDFSUDVAYNRTRGK-----DTDTGEYISSINPDT	568
QY	650	AQDDQNAAPRVPAAERLGHPLKASLTDIRIDANDLYYRVFAQNKLARYETRTPGHMHMLNCAN	709
DB	569	VTSTLN--IPIAHSGF-----SVGWVGITFADRS-----THSSS	600
QY	710	YRNRTRYGENWNWY	723
DB	601	YSKPQPGYGVNDFTV	614

```

RESULT 38
AC3334
metal chelate outer membrane receptor [imported] - Brucella melitensis (strain 16M)
C/Species: Brucella melitensis
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 23-Dec-2002
C/Accession: AC3334
R/DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

```

A;Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A;Reference number: AD3252; PMID:11756688
A;Accession: AC3334
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-599 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL51838.1; PID:gl7982585; GSFDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BME10657
A;Map position: I
C;Superfamily: vitamin B12 receptor; tonB-dependent receptor amino-terminal homology; to

Query Match 4.6%; Score 187.5; DB 2; Length 599;
Best Local Similarity 21.7%; Pred. No. 2.6e-05;
Matches 151; Conservative 72; Mismatches 226; Indels 247; Gaps 37;

QY 26 ETBQSQGLEVTVVVVGKSRPRATSGLLHTSTASDKII-SGDTLRQKAVNIGDALDGVPGI 84
DB 7 DKXDGVTLDIIVT-----PLRRASSLQRTSSVSVTDADIERSAAPDQSLQLQTVSGIS 62

QY 85 ASQYGG-GASAPV-IRQGTGRRIKVLNHHGETGDMADFSPDHAIMVDITALSQOVELRGP 142
DB 63 VKTNGQGGSSADIVMRGSSKQIWLNVGRTASATSGSTALANIPITSI-ERIEIARGA 121

QY 143 VTLLYSSGNVAGLDV---ADGIKEP-----MEENGVSGLGLRLSGNLEKUTSGGINI 195
DB 122 HSSQYGADATGGVINIITKGGACGERAWCGSVSTGVSHPWG-GYASGSLQGRSSDGIDY 180

QY 196 GLGKNFVLHTEGLYKRGSDYAVPYRNLKELPSPRRFANGQHRVIG---WRKEFYRR 251
DB 181 AVGAFT-GRQY-----DFTTP--EAFGHEPDDD-GFLGGSFNFAKSKDFDWGKIYADG 231

QY 252 TYSDRDQYGLPAHS-HEYD-----DCHADIIWQKSLINKRYLQLYPHLLTEE 298
DB 232 LFSGRGNQYDAPAFNEADSTAGTKVGNRIDHTAD--WSSTV-----EF 275

QY 299 DVYDNPGLSCGFHDDDDAHAAHNGKPTDILNKRYELRAEWKQPPP-----GF 348
DB 276 STGIDN---SRNFRKGIE-----GSDWPETR--RYGVFEASTEKGSDTKGVSHVVTVGGV 323

QY 349 EALRVHLNRN-DVHHDEKAGDAVENFPNNQTNQRIELRHQPIRLKGSNGVOYLQKSS 407
DB 324 EAYREKINTIDY--DETGRDLAVP-----GQVSLEY-----354

QY 408 ALSATSEAVKQPMLDNKNVQHYGFFGVEQANWDFNTEGGVRVEKOKASIRYD-KALIDR 466
DB 355 -----DALRFDGG-----IRYDHNQFGN 373

QY 467 ENYVK-----QPLPDLCAGHQRTARSFALSGNWTFTPQHKLSLTASHQERLPSTQELIYAHG 521
DB 374 VTTYNIGASYEILPDLVLRSSYATGF-----BAPTNELLYYFG 411

QY 522 KHVATNTFVGNGKHLNKKERSNNIELAGYBGDRWQ-----YNLALYRNFNGYI-----570
DB 412 -----FANPDLQPEKRSRVEGL-----NQWATSTSLDMALYQTRLSDAIMSTAP 457

QY 571 -----YAQTLNDRGPKSIEDSDMKLVRYNQSCADF-----YCAEG 607
DB 458 SYIPYNIASAKVTGLEATLHSHSFNEQWIKGMVD---LKEPVEDSGNDLPYRERFKAA 514

QY 608 EIVEKP-----TPRIRIGVS-----GDYV 626
DB 515 EVNFKPVEKLDLTVRLVYGGSSRYTNKTKGLGDYV 550

RESULT 39
ORSCIC
colicin I receptor precursor - *Escherichia coli* (strain K-12)
C;Species: *Escherichia coli*
C;Date: 30-Sep-1990 #sequence revision 05-Dec-1997 #text change 01-Mar-2002
C;Accession: B64984; A33056; A33868; A28377; C41871; A35408; S24561
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B64984
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-663 <BIAT>
A;Cross-references: GB:AE000304; GB:U00096; NID:g1788470; PIDN:AAC75216.1; PID:g1788478
A;Experimental source: strain K-12, substrain MG1655
R:Nau, C.D.; Konisky, J.
J. Bacteriol. 171, 1041-1047, 1989
A;Title: Evolutionary relationship between the TonB-dependent outer membrane transport
A;Reference number: A32056; MUID:89123100; PMID:2644220
A;Accession: A32056
A;Molecule type: DNA
A;Residues: 1-146, 'RCARCSEVHQ', 159, 'NRSEMV', 166, 'YRYRYHSGTSRSR', 182-527, 'N', 529-613
A;Note: the authors translated the codon AAT for residue 528 as Ile
R:Nau, C.D.; Konisky, J.
J. Bacteriol. 171, 4530, 1989
A;Reference number: A33868
A;Contents: corrections
A;Accession: A33868
A;Molecule type: DNA
A;Residues: 144-184;610-663 <NA2>
R:Griggs, D.W.; Tharp, B.B.; Konisky, J.
J. Bacteriol. 169, 5343-5352, 1987
A;Title: Cloning and promoter identification of the iron-regulated cir gene of Escherichia coli
A;Reference number: A28377; MUID:88058737; PMID:3316180
A;Accession: A28377
A;Molecule type: DNA
A;Residues: 1-59 <GR1>
A;Cross-references: GB:M19295; NID:g145545; PIDN:AAA23581.1; PID:g551794
A;Note: residues 26-45 were confirmed by protein sequencing
R:Steffes, C.; Ellis, J.; Wu, J.; Rosen, B.P.
J. Bacteriol. 174, 3242-3249, 1992
A;Title: The lysE gene encodes the lysine-specific permease.
A;Reference number: A41871; MUID:92250419; PMID:1315732
A;Accession: C41871
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-96, 'D', 98-125 <STE>
R:Griggs, D.W.; Kafka, K.; Nau, C.D.; Konisky, J.
J. Bacteriol. 172, 3529-3533, 1990
A;Title: Activation of expression of the Escherichia coli cir gene by an iron-independent
A;Reference number: A35408; MUID:90264362; PMID:2160948
A;Accession: A35408
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-21 <GR2>
C;Comment: This outer membrane receptor for colicins Ia and Ib is regulated by both celI and celII genes.
C;Genetics:
A;Gene: cir; cirA; feuA
A;Map position: 43 min
C;Superfamily: ferriterochelin receptor; tonB-dependent receptor amino-terminal homol
C;Keywords: iron transport; membrane protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-624/Product: colicin I receptor #status experimental <MAT>
F;66-211/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F;365-663/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 4.6%; Score 187.5; DB 1; Length 663;
Best Local Similarity 20.0%; Pred.No. 3.1e-05;
Matches 172; Conservative 110; Mismatches 271; Indels 307; Gaps 42;

Qy 6 LKEPIV-----LSLLINTPLLAQHETEQVGSLFTVTWVKSRPRATSGLLHTSTASDKI 60
Db 4 LNPFVRVGLCLSAISCAMPVLAVDDDG-----ETWVTASS-----VEQNKLKPAPISIV 53
Qy 61 ISGDTLRQXAV-NLGDALGVPGIHASQVGGGASAPVIRG-----QTGRIK---- 106

Db 54 ITQEDLQKRPVQNLKDVLEKVEPGVQLTNEGNRKGVSIRGLDSSVTLILVDGKRVNSRNA 113

Qy 107 VLNHGEGTGMADSPDHAIMVDTSALSOQVELRGPVTLTYSSGNVAGLVADQKIEK 166

Db 114 VFRN-----DFDLW-IPVDSI--ERIEVRGPMSSILGSDALGVNMIITKIQCK 163

Qy 167 MPENGVSGLGL-----RLSGNLEKLTSGGINTGLGNFVLHTTEGLYRKSGDYAV 217

Db 164 W-----SGTVVDTTIOEHRDRGDTYNGQFFTSGLIDGV-----LGM 201

Qy 218 PRYRNL-KELPDSPRFA--NGQRAVLGWRKRYRYTSDRRDQYG-LPAHSHEY--- 269

Db 202 KAYGLAKREKDDPONSTTTTGTGETPRIEGFSR-----DGNVEFAWTEPNQNHDTAG 254

Qy 270 -----DCHADLIWCKSLINKRY-----LQLYPHLLTEEDVDVDPNGLSCG 310

Db 255 YGFDEQDRDSLSLKNRLERQNYSVSHNGRDYGTSELKY-----GKVENKPNPNSP 309

Qy 311 FHDDDDAHAHNGKFWIDL--NKRYELRAEWKOPFPFGFEALRVHLNRNDYHHDEKAGD 368

Db 310 I-----TSESNTVDGKYTLPLTAINGFLTVGGEW-----HD-KLSD 345

Qy 369 AVENFFNQTQARIELRHQPIRLGLKSGWGYLGGKSSALSATSEAVKQPMLLDNKYOH 428

Db 346 AV-----NLT-----GTSKTSAS-----QY 362

Qy 429 YSFFGEOANMNDFTLEGGVRVEKOKASIRYDKALIDRENTYKQPLDLAGHRQTARSA 488

Db 363 ALFVEDEWRIFEPLATTGVMD-----DHETGEHWSPRAYLVYNATDVT 409

Qy 489 LSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVANTTE-----VGNKHLNKRNN 543

Db 410 VKGW-----ATAF---KAPSLQL---SPDWTNSCRGACKIVGSPDLKPETSES 454

Qy 544 IELALYGBGR-W-----OYNALAYRN-----RFGNVIYAQTLNDGRG 580

Db 455 WELGLYNGEGLWEGVSVTVFNDVKDRISISRTSDVNAAPGYQNFVGFETGANGRR 514

Qy 581 PKSIEDDEMKLVRYNQSGADPYGAEIYFKTPRYRIGVSGDYVRGLKNLPLSGRE 640

Db 515 IPVFS-----YNNVKARIQGVETELKIPFNDEWKLSINYTYNDGR----- 555

Qy 641 DAYGNRPFIADQDQAPRPAARLGFHLKASITDRIDANLDYRVF-----AQKLARY 694

Db 556 -----DVSNGENKPLSDLPFHTANGTLDWKPLALEDWSVSGHYTGQKRADS 604

Qy 695 ETRTPGHMLNLGANYRNTRYGEWN---WYVKAD-----NL-----LNQSVYAHSS 738

Db 605 TAKTPGGTYI-----WNTGNAWQVTKDKVLKAGVLNLGDKLSDRDDYSYN- 649

Qy 739 FLSDTQMGRSFTGGVNVKF 758

Db 650 -----EDGRRYFMAVDYRF 663

RESULT 40

AB0124

probable TonB-dependent outer membrane receptor YPO1011 [imported] - Yersinia pestis (st

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AB0124

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Li, M.; Rutherford, K.; Skellern, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AB0124

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-690 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC89853.1; PID:g15979079; GSPDB:GN00175

C:Genetics:

A:Gene: YPO1011

Query Match 4.6%; Score 187.5; DB 2; Length 690;

Best Local Similarity 20.3%; Pred. No. 3.3e-05;

Matches 189; Conservative 127; Mismatches 299; Indels 241; Gaps 43;

Qy 14 LLINTPLLA--QAHETEQSGLVETVTVGKSRPRATSLHTSTASDKIISG-DTLRQKA 70

Db 13 VLISLPCLAWSQSHNSKD-ELDITVVAQK-----INQQOKTPTISIVLTGFDLERENI 67

Qy 71 VNLGDALDVGPIIASQYGGGASAP--VIRGOTGRRIKVLNHHGETG-DMADFSPP--HA 125

Db 68 ENIVESIMRIENVMVMKAGNPSDAGFTMRGTT-----PGMEGIVSQSGFFIDGVYA 118

Qy 126 IMVDTAL--SQQVELRGPVTLTYSSGNVAGLVADQKIPKOMPENGVSGLGLRSLSSG 183

Db 119 NTFDELLDVRIEVLRFQATLYGRNTEGSGVINVT-KDPEFSPEY---KIGUSYGY 173

Qy 184 NLEKITS---GGINIGLGKNF-----VLHTEGLYRKSGDYAVPRYRNLKRLPDSPPRFAN 235

Db 174 NRTQVTVTLGGSINDSEQFSYRAALKYLYGNGYFKRDYD---GKNVNDNLNDFSGRPKL 229

Qy 236 GQRAVLGWRKRFRTYSDRRDQYGLPAHS-----HEYDDCH--ADIIWQSLI 283

Db 230 RWQPMDDGMD--VMTTFDIQNRNGNTSFTALDKIKSGQKYVDSNYIGKSDVDAYKGQV 286

Qy 284 NKRYLQLYPHLLTEEDVDVDPNGLSCGFHDDDDAHAAHNGKFWTD----- 329

Db 287 NAVY-----TFDDIDFTS-----VSAYVDERKVDNQDLDTRLSISEL 324

Qy 330 LRNKR-----YELRAEWKOPFPFGFEALRVHLNRNDYHHDEKAGDAVENFFNQTQAR 382

Db 325 LMRKTKQFSQEFRLNSKYSGFFNWL-----IGSYFYQDDENE 363

Qy 383 IELRHQP--IGRLKSGWGVYLGQSSALSATSEAVKQPMLLDNKYHSPFG-VEQANW 439

Db 364 IDFRYLPYLAQLRKS-----DITNNYAVFGVNYNYLL 397

Qy 440 DNFTLEGGVRVEKOKASIRYDKALIDRENTYKQPLDLAGHRQTARSAFA---LSGNWY 494

Db 398 NDVELVAGARYDYKELNLF---LMDNGFNYPQY-----SHDNNNSFGAFPLKVLNYY 450

Qy 495 FTPQHLK--SLTASHQERLPST--QELYAHGKHVATNTFEVGNKHLNKRNNIELALG 549

Db 451 ITGDAMLYTSIARGYKSGGFNTLGPQSSRAYNAEYMT-TYEAGVKTWFDRIT- 501

Qy 550 YEGDRWQYNLALYRNFQNYIYAQTLNDGRGPKSIEDDEMKLRYNOSGADFYGAEHI 609

Db 502 ---VRWNTSL-----FWNDMKDQOQVEVAYYPIYSYVNSGKSLSR-----GLESEL 543

Qy 610 YFKPTPRYRIGVSGDYVRGLKNLPLSGREDAY-----GNRPFIADQDQAPRPAAR 663

Db 544 AWRIITRGLTVSANVGYTDAFYKNPTEIKVDNNYIPVNYKGNRP-----ANSP----- 591

Qy 664 LGFHLKASLTDRIIDANLDYRVFPAQNKLARVET-----RTPGHMLNLGANYR 711

Db 592 -----GYTYSIGADYNFLNGYFVN--ATYNVKGSTYLDNANSKKQPAYGLDLDTAGY- 641

Qy 712 RNTRYGEMWVVKADNLLNQSVYAHSSFLSTPQWGR-----SFTGGVNVKF 758

Db 642 ENKDYGV-NWIK--NILDETYVTRAFKMDDDGIWYGRAGEPINFQGVNFNVKF 690

Search completed: November 14, 2003, 11:00:39

Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:01:18 ; Search time 17 Seconds
(without alignments)
2096.838 Million cell updates/sec

Title: US-09-936-377-2
Perfect score: 758
Sequence: 1 MAQTTLPLIVLSILLINTEPL.....FLSDTPQMGRSFTGGVNVKF 758

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_41.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.2	330	1 PE42 ARATH	Q95b81 arabidopsis
2	9	1.2	1980	1 MY9B RAT	Q63358 rattus norv
3	9	1.2	2114	1 MY9B MOUSE	Q93y06 mus musculu
4	9	1.2	2158	1 MY9B HUMAN	Q13459 homo sapien
5	8	1.1	196	1 YO96 AGRT5	P58792 agrobacteri
6	8	1.1	316	1 BLA3 BACCE	P06548 bacillus ce
7	8	1.1	400	1 FDH PSESR	P33160 pseudomonas
8	8	1.1	933	1 RGA4 SCHPO	O74360 schizosacch
9	8	1.1	1009	1 RGA2 YEAST	Q06407 saccharomyc
10	7	0.9	117	1 NUSM DROME	P19930 drosophila
11	7	0.9	117	1 NUSM DROSA	P51940 drosophila
12	7	0.9	117	1 NUSM DROXA	P07705 drosophila
13	7	0.9	152	1 PYRI ECOLI	P00478 escherichia
14	7	0.9	152	1 PYRI SALT1	Q82130 salmonella
15	7	0.9	152	1 PYRI SALT2	P08421 salmonella
16	7	0.9	161	1 Y004 BPL2	Q42539 bacterioph
17	7	0.9	200	1 ATR4 IPOBA	Q40089 ipomoea bat
18	7	0.9	212	1 DEOC HALNI	Q9np08 halobacteri
19	7	0.9	225	1 Y116 METJA	Q57580 methanococc
20	7	0.9	230	1 FA3A HUMAN	P98173 homo sapien
21	7	0.9	230	1 FA3A MOUSE	Q948t0 mus musculu
22	7	0.9	231	1 KCY STRCO	Q9eww6 streptomyce
23	7	0.9	249	1 ULA3 HCVIA	P16734 human cytom
24	7	0.9	281	1 YE94 XILPA	Q9pd85 xyliella fas
25	7	0.9	308	1 HEM4 RHIL0	Q98ei7 rhizobium l
26	7	0.9	315	1 TPIC ARATH	Q9skp6 arabidopsis
27	7	0.9	334	1 FLGJ VIBCH	Q9kg15 vibrio chol
28	7	0.9	335	1 XYNB STRLI	P26515 streptomyce
29	7	0.9	349	1 TRM1 ARCFU	Q29443 archaeoglob
30	7	0.9	368	1 MLTA VIBCH	Q9kpq4 vibrio chol
31	7	0.9	389	1 HEM2 PEA	Q43082 pisum sativ
32	7	0.9	388	1 ALR MYCLE	P38056 mycobacteri
33	7	0.9	388	1 PEPF RABIT	P27823 oryctolagus

Q24093	drosophila	1	A23D DROME	398	0.9	7	34
P05783	homo sapien	1	K1CR_HUMAN	429	0.9	7	35
P34693	caenorhabdi	1	SYT1_CABEL	441	0.9	7	36
P33705	canis famli	1	CD4_CANFA	463	0.9	7	37
P47867	mus musculu	1	SG3_MOUSE	471	0.9	7	38
P47868	rattus norv	1	SG3_RAT	471	0.9	7	39
P47455	mycoplasma	1	Y213_MYCGE	471	0.9	7	40
O35280	mus musculu	1	CHK1_MOUSE	476	0.9	7	41
P53789	oryctolagus	1	VTDB_RABIT	476	0.9	7	42
O10439	mouse adeno	1	PEN3_ADEMI	489	0.9	7	43
P11636	neurospora	1	QAY_NEUCR	537	0.9	7	44
P38346	saccharomyc	1	YB90_YEAST	545	0.9	7	45
O67622	aquifex aeo	1	YH32_AQUAE	558	0.9	7	46
Q924u5	rattus norv	1	TES2_RAT	570	0.9	7	47
Q9h1r3	homo sapien	1	KML2_HUMAN	595	0.9	7	48
P07313	oryctolagus	1	KML2_RABIT	607	0.9	7	49
P43634	saccharomyc	1	CHM4_YEAST	648	0.9	7	50
P26804	friend mufi	1	ENV_MLVFF	676	0.9	7	51
Q17390	caenorhabdi	1	CUL2_CABEL	743	0.9	7	52
Q91cc2	anabaena ep	1	PHYA_ANASP	765	0.9	7	53
Q92G16	listeria in	1	SYFB_LISIN	802	0.9	7	54
Q8Y7q1	listeria mo	1	SYFB_LISMO	802	0.9	7	55
Q92i38	rickettsia	1	SYFB_RICCN	818	0.9	7	56
P22174	infectious	1	RRPO_IPNVS	844	0.9	7	57
P53532	corynebacte	1	CLPB_CORGL	852	0.9	7	58
Q9e633	mus musculu	1	UBPT_MOUSE	869	0.9	7	59
Q8k9u3	buchnera ap	1	SECA_BUCAP	874	0.9	7	60
Q97ig3	clostridium	1	SYA_CLOAB	881	0.9	7	61
Q01454	saccharomyc	1	CTF4_YEAST	926	0.9	7	62
Q01168	magnaporthe	1	NUT1_MAGGR	956	0.9	7	63
P39083	saccharomyc	1	RGAL_YEAST	1007	0.9	7	64
P55131	actinobacil	1	RT32_ACTPL	1052	0.9	7	65
Q91783	xenopus lae	1	EG52_XENLA	1067	0.9	7	66
Q9vys3	drosophila	1	RNT1_DROME	1180	0.9	7	67
Q10435	schizosacch	1	YDE1_SCHPO	1647	0.9	7	68
O84337	chlamydia t	1	UVRA_CHLTR	1786	0.9	7	69
Q43451	homo sapien	1	MGA_HUMAN	1856	0.9	7	70
P46939	homo sapien	1	UTRO_HUMAN	3433	0.9	7	71
P98161	homo sapien	1	PKD1_HUMAN	4303	0.9	7	72
Q39575	chlamydomon	1	DYHG_CHLRE	4485	0.9	7	73
Q96pk2	homo sapien	1	MAC4_HUMAN	5938	0.9	7	74
P13948	escherichia	1	PIL3_ECOLI	21	0.8	6	75
P50370	chlamydomon	1	YC12_CHLRE	33	0.8	6	76
P39495	bacterioph	1	Y12C_BPT4	61	0.8	6	77
P50093	homo sapien	1	CU99_HUMAN	68	0.8	6	78
O78442	guillardi	1	POZ2_GUITH	76	0.8	6	79
Q9xsu8	ralstonia s	1	ACP2_RALSO	81	0.8	6	80
P03286	human adeno	1	Y95_ADE02	95	0.8	6	81
Q9fnc9	arabidopsis	1	OM22_ARATH	99	0.8	6	82
Q8m9t9	chaetosphe	1	NULC_CHAGL	100	0.8	6	83
P81762	pleurotus o	1	RNPO_PLEOS	101	0.8	6	84
P26558	human papil	1	VE7_HPV51	101	0.8	6	85
P20884	human immun	1	VFR_HV2ST	104	0.8	6	86
Q97484	euplotus cr	1	H2B_EUPCR	113	0.8	6	87
Q95881	peromyscus	1	NU3M_PERGO	115	0.8	6	88
Q9slf7	arabidopsis	1	RA2A_ARATH	115	0.8	6	89
P24957	pisaster oc	1	NU3M_PISOC	116	0.8	6	90
P48941	chondrus cr	1	RT11_CHOCR	116	0.8	6	91
Q36122	methanobact	1	RL24_METH	117	0.8	6	92
Q9z9h5	thermus the	1	RL17_THETH	118	0.8	6	93
P14494	escherichia	1	PIL5_ECOLI	119	0.8	6	94
P14495	escherichia	1	PIL6_ECOLI	119	0.8	6	95
P14496	escherichia	1	PIL7_ECOLI	119	0.8	6	96
P10513	escherichia	1	PIL4_ECOLI	120	0.8	6	97
P04737	escherichia	1	PIL1_ECOLI	121	0.8	6	98
P01169	lophius ame	1	SMS1_LOPAM	121	0.8	6	99
P17792	agrobacteri	1	VIB2_AGRTS	121	0.8	6	100
P43575	saccharomyc	1	PAU5_YEAST	122	0.8	6	101
P07480	sus scrofa	1	GALA_PIG	123	0.8	6	102
Q12370	saccharomyc	1	YL25_YEAST	124	0.8	6	103
P78946	schizosacch	1	RL26_SCHPO	126	0.8	6	104
Q11035	mycobacteri	1	YD65_MYCTU	128	0.8	6	105
P58349	rhizobium m	1	RBS1_RHIME	129	0.8	6	106

107	1	RBS2 RHIME	129	0.8	6	180	199	1	PSBY SPIOL	180470 spinacia ol
108	1	CRCE METKA	131	0.8	6	181	201	1	ARP_RAT	Q63055 rattus norv
109	1	VG40_HSV11	136	0.8	6	182	201	1	RECR_UREPA	Q9pr56 ureaplasma
110	1	Y877_ARCFU	137	0.8	6	183	202	1	VNS1_IAPOM	Q3499 influenza a
111	1	V82_SOLIN	138	0.8	6	184	202	1	Y200_PVRAB	Q9v274 pyrococcus
112	1	YQBS_BACSU	141	0.8	6	185	203	1	RYL1_VARLI	P41924 yarrowia li
113	1	OM22_HUMAN	142	0.8	6	186	204	1	MDCG_PSEPU	Q9z448 pseudomonas
114	1	DTD_BACHD	146	0.8	6	187	204	1	Y275_METUA	Q57723 methanococcus
115	1	MUCA_SALTY	146	0.8	6	188	208	1	GTP1_CABEL	P10299 caenorhabdi
116	1	YD82_METUA	146	0.8	6	189	208	1	RS6_MYCCE	Q9pfs7 xylella fas
117	1	YDPK_ECOLI	147	0.8	6	190	208	1	KSH_XYLFA	P47336 mycoplasma
118	1	MR_PPYPO	147	0.8	6	191	209	1	PCX8_ACICA	P20371 acinetobact
119	1	Y236_METUA	147	0.8	6	192	210	1	ALXH_TREPA	Q83578 t putative
120	1	Y246_HAEIN	148	0.8	6	193	210	1	HIS2_VIBCH	Q9ksw7 vibrio chol
121	1	DTD_GLOAB	149	0.8	6	194	211	1	YD71_MYCPN	Q9ru32 deinococcus
122	1	3DHQ_EMENI	153	0.8	6	195	211	1	CCMA_HAEIN	P75410 mycoplasma
123	1	FLIL_ECOLI	154	0.8	6	196	212	1	KAD_STRAP3	P54032 haemophilus
124	1	FLIL_SALTY	154	0.8	6	197	212	1	KAD_STRAP3	Q8e780 streptococc
125	1	IL2_MERUN	155	0.8	6	198	212	1	KAD_STRAP3	Q8k8x1 streptococc
126	1	V17K_BSMV	155	0.8	6	199	212	1	KAD_STRAP3	Q8p24 streptococc
127	1	YBEY_ECO57	155	0.8	6	200	212	1	KAD_STRAP3	Q8p24 streptococc
128	1	YBEY_ECOL6	155	0.8	6	201	212	1	Y434_RICPR	Q05865 rickettsia
129	1	YBEY_ECOLI	155	0.8	6	202	212	1	YGBL_ECOLI	Q46890 escherichia
130	1	URBE_CORGL	157	0.8	6	203	212	1	YJ47_ARCFU	O28332 archaeoglob
131	1	URBE_KLEAE	158	0.8	6	204	213	1	MDCG_XANAC	Q8ppw9 xanthomonas
132	1	NURD_ECO57	159	0.8	6	205	214	1	EPD_NOTCH	Q9t127 cyanidium c
133	1	NURD_ECOLI	159	0.8	6	206	215	1	CYB5_CVACA	Q9t127 cyanidium c
134	1	YQGF_RHOCA	159	0.8	6	207	215	1	EPD_CARAU	P13506 carassius a
135	1	COAT_PMWV	160	0.8	6	208	215	1	EPD_CARAU	P12958 carassius a
136	1	COAT_TMOB	160	0.8	6	209	215	1	EPD_CYPCA	P38528 cyprinus ca
137	1	RISB_SYNCTU	160	0.8	6	210	215	1	ERD2_ARATH	P35402 arabidopsis
138	1	YF85_SYNEL	160	0.8	6	211	215	1	ERD2_PETHY	Q9zt22 petunia hyb
139	1	FURI_MYCFO	161	0.8	6	212	215	1	ORN_MYCLE	O07708 mycobacteri
140	1	PTTB_ECOLI	161	0.8	6	213	215	1	ORN_MYCTU	O06174 mycobacteri
141	1	YD22_HAEIN	161	0.8	6	214	216	1	YQW8_YEAST	P53084 saccharomyc
142	1	YP49_RALSO	161	0.8	6	215	217	1	DEOC_MYCHO	P43048 mycoplasma
143	1	DH1B_ORISA	164	0.8	6	216	217	1	EPD_BRARE	P17561 brachydanio
144	1	DH1C_ORISA	164	0.8	6	217	217	1	FDNI_ECOLI	P24185 escherichia
145	1	RISB_SYN3	164	0.8	6	218	217	1	VNS1_IANAN	P21431 influenza a
146	1	YPOC_BACSU	164	0.8	6	219	217	1	VNS1_IACKO	O57268 influenza a
147	1	FIB_STRAM	165	0.8	6	220	218	1	EPD_DANAE	Q30399 danio asqu
148	1	YJGM_SALTY	165	0.8	6	221	218	1	SGAR_MYCPN	P75293 mycoplasma
149	1	YPOE_BACSU	168	0.8	6	222	218	1	TRPF_BACHD	Q9KCB1 bacillus ha
150	1	DEF_XYLFA	170	0.8	6	223	218	1	Y4VH_RHISN	Q53216 thizobium s
151	1	AR21_SCHPO	174	0.8	6	224	221	1	FGF3_CHICK	P48801 gallus gall
152	1	CYGB_BRARE	174	0.8	6	225	221	1	YGB1_HAEIN	P44290 haemophilus
153	1	FMA_SERMA	174	0.8	6	226	222	1	FLA5_PVRAB	Q9uy14 pyrococcus
154	1	Y201_AQUAE	174	0.8	6	227	222	1	NANE_STRAM	Q99w8 staphylococ
155	1	CRIZ_PANAN	175	0.8	6	228	222	1	NANE_STRAM	Q8ny4 staphylococ
156	1	HSLV_AQUAE	176	0.8	6	229	226	1	CG5_MOUSE	Q9da57 mus musculu
157	1	ATPQ_DROME	178	0.8	6	230	227	1	TPIS_SULSO	Q97vm8 sulfolobus
158	1	PYRE_ARCFU	178	0.8	6	231	227	1	VNS1_IACKJ	P08274 influenza a
159	1	PYRR_CLOPE	178	0.8	6	232	227	1	VNS1_IADU3	P08272 influenza a
160	1	DEF2_ANASP	179	0.8	6	233	227	1	VNS1_IAMU3	P08270 influenza a
161	1	YFAZ_ECOLI	180	0.8	6	234	227	1	VNS1_IATFS	P08268 influenza a
162	1	YPKW_THECU	180	0.8	6	235	227	1	VNS1_IATFS	P08278 influenza a
163	1	APT_CAMEB	182	0.8	6	236	228	1	TRMD_UREPA	Q9p22 ureaplasma
164	1	DEF_STRAM	183	0.8	6	237	230	1	VNE1_CVBN	P10526 bovine coro
165	1	PYRE_METCA	183	0.8	6	238	230	1	VNE1_CVHOC	Q01455 human coron
166	1	TFA_BPT4	183	0.8	6	239	230	1	VNE1_CVTKE	P26021 turkey ente
167	1	YCF4_CHLVU	183	0.8	6	240	230	1	VNS1_IACKB	P36349 influenza a
168	1	SYRP_LACHI	184	0.8	6	241	230	1	VNS1_IACKB	P36349 influenza a
169	1	APT_HALNI	188	0.8	6	242	230	1	VNS1_IALFI	P31500 influenza a
170	1	ATKE_SYN3	190	0.8	6	243	230	1	VNS1_IALFI	P26131 influenza a
171	1	PPIA_ERWCH	190	0.8	6	244	230	1	VNS1_IALFI	P17042 influenza a
172	1	YK27_VIBCH	193	0.8	6	245	230	1	VNS1_IAMAO	P13138 influenza a
173	1	YCDI_ECO57	196	0.8	6	246	230	1	VNS1_IAP10	P13140 influenza a
174	1	YCDI_ECOLI	196	0.8	6	247	230	1	VNS1_IAP10	P13142 influenza a
175	1	DSBE_XYLFA	197	0.8	6	248	230	1	VNS1_IAP10	P13142 influenza a
176	1	GCH2_NEIMA	197	0.8	6	249	231	1	HIS1_RHIME	Q92r16 thizobium m
177	1	GCH2_NEIMB	197	0.8	6	250	232	1	GIDB_COREF	Q8rs1 corynebacte
178	1	HPAP_RALSO	197	0.8	6	251	232	1	HIS1_RHIL0	Q98758 thizobium l
179	1	RS5_ARCFU	198	0.8	6	252	232	1	VATE_TREPA	O83439 treponema p

253	6	0.8	233	1	MENH_BACSU	P31113	6	0.8	293	1	HSLO_STAAM	Q99w91	staphylococ
254	6	0.8	236	1	ABME_RABIT	P47855	6	0.8	293	1	HSLO_STAAM	Q8nx23	staphylococ
255	6	0.8	237	1	LIVF_ECOLI	P22731	6	0.8	293	1	YEIE_ECOLI	P32484	escherichia
256	6	0.8	237	1	VNSI_IAAIC	Q02599	6	0.8	294	1	DAPA_THEME	Q8x1k9	thermotoga
257	6	0.8	237	1	VNSI_IACHI	P11618	6	0.8	294	1	MALK_ENTAE	P18813	enterobacte
258	6	0.8	237	1	VNSI_IAPOF	P03497	6	0.8	294	1	RR2_FORPU	P51311	porphyra pu
259	6	0.8	237	1	VNSI_IAUSS	P03498	6	0.8	295	1	DAPA_OCEIH	Q8eg11	oceanobacil
260	6	0.8	238	1	RL2_AERPE	Q9vfn1	6	0.8	295	1	HTPX_LEPIN	Q8exn4	leptospira
261	6	0.8	238	1	YJHA_ECOLI	P39372	6	0.8	295	1	ISPE_XANCP	Q8pc64	xanthomonas
262	6	0.8	240	1	VP26_NPVAC	P08358	6	0.8	296	1	MOAA_ARCFU	Q28273	archaeoglob
263	6	0.8	244	1	APMI_HUMAN	Q15848	6	0.8	296	1	PCOB_ECOLI	Q47453	escherichia
264	6	0.8	244	1	OPAI_NEIGO	Q04877	6	0.8	298	1	TPIC_SECCB	P46223	secale cere
265	6	0.8	246	1	YIAT_ECO57	P58224	6	0.8	298	1	YIHV_ECOLI	P32143	escherichia
266	6	0.8	246	1	YIAT_ECOLI	P37681	6	0.8	299	1	HTPX_STRPN	Q79qdc	streptococc
267	6	0.8	247	1	APMI_MOUSE	Q60994	6	0.8	299	1	HTPX_STRR6	Q8dph5	streptococc
268	6	0.8	249	1	KDKA_XANAC	Q8phb3	6	0.8	300	1	RIP3_MAIZE	P25891	zea mays (m
269	6	0.8	249	1	TL3X_MOUSE	Q9et35	6	0.8	301	1	FLIP_RHOSH	Q85133	rhodobacter
270	6	0.8	251	1	HEMA_PSEAE	P48246	6	0.8	302	1	FDHE_SHEON	P28522	zea mays (m
271	6	0.8	252	1	PRGK_SALTY	P41786	6	0.8	302	1	YET6_YEAST	P59188	shewanella
272	6	0.8	253	1	PYHD_CVOP	P36701	6	0.8	302	1	CHEV_BACSU	P40065	saccharomyc
273	6	0.8	255	1	HA21_MOUSE	Q36835	6	0.8	303	1	RIP9_MAIZE	P37599	bacillus su
274	6	0.8	255	1	HA22_MOUSE	P01904	6	0.8	304	1	Y007_MYCTU	P25892	zea mays (m
275	6	0.8	255	1	RS2_THETH	P04224	6	0.8	304	1	NAC_ECOLI	P71575	mycobacteri
276	6	0.8	255	1	Y166_METJA	P80371	6	0.8	305	1	NAC_KLEAE	Q47005	escherichia
277	6	0.8	255	1	ATP6_HANWI	Q57630	6	0.8	305	1	VP53_BPAPS	Q86597	klebsiella
278	6	0.8	256	1	GRST_BACBR	P48879	6	0.8	307	1	FLGJ_VIBPA	Q9t1p5	bacterioph
279	6	0.8	256	1	SURE_AGR5	P14686	6	0.8	308	1	MRAW_UREPA	Q9x9j3	vibrio para
280	6	0.8	256	1	RSEA_BACSU	Q8ueg3	6	0.8	308	1	Y117_HELPJ	Q9pqa4	ureaplasma
281	6	0.8	258	1	RFBA_MYXKA	P39650	6	0.8	308	1	Y117_HELPJ	Q9zmv6	helicobacte
282	6	0.8	260	1	Y166_METJA	Q50862	6	0.8	308	1	Y117_HELPJ	P56080	helicobacte
283	6	0.8	261	1	YZ78_PSEAE	P03256	6	0.8	309	1	YJ09_AQUAE	P74422	synecocyst
284	6	0.8	261	1	Y4JP_RHISN	Q9hy42	6	0.8	309	1	YJ09_AQUAE	Q67744	aquifex aeo
285	6	0.8	262	1	Y4JP_RHISN	P55516	6	0.8	309	1	YJ48_STROO	Q9x8h1	streptomyce
286	6	0.8	265	1	DH47_ARATH	P31168	6	0.8	310	1	RYAD_ECOLI	P17963	escherichia
287	6	0.8	266	1	NAGB_VIBCH	Q9kks5	6	0.8	311	1	ESCA_ARATH	Q987c9	arabidopsis
288	6	0.8	266	1	NAGB_VIBCH	Q8d4t9	6	0.8	312	1	CELF_PRVKA	Q85230	pseudorabie
289	6	0.8	266	1	DHPS_STAAU	P17166	6	0.8	312	1	TRXB_CHLTR	Q84101	chlamydia t
290	6	0.8	267	1	NADP_PSEAE	Q05701	6	0.8	312	1	VR2B_EPT4	P03691	bacterioph
291	6	0.8	267	1	NADP_PSEAE	Q8zx14	6	0.8	312	1	Y353_EUCAP	Q8k912	buchnera ap
292	6	0.8	267	1	RL4_SULSO	Q9uxa6	6	0.8	313	1	CBR1_RHIME	P58332	rhizobium m
293	6	0.8	267	1	SV72_ARATH	Q94kk6	6	0.8	313	1	CBR2_RHIME	P56885	rhizobium m
294	6	0.8	268	1	HLVC_TREHY	Q54318	6	0.8	313	1	PLDB_HABIN	P44800	haemophilus
295	6	0.8	268	1	SUHB_AERPE	P14628	6	0.8	314	1	CSB_SULSO	P58030	sulfolobus
296	6	0.8	268	1	SUHB_AERPE	Q9yaz7	6	0.8	314	1	TPIC_FRAAN	Q9m4s8	fragaria an
297	6	0.8	269	1	DAPP_CHLMO	Q9pjw2	6	0.8	317	1	LDH_CLOPE	Q8xp62	clostridium
298	6	0.8	269	1	PHYD_CPVES	P36326	6	0.8	318	1	EX53_MYCTU	Q10699	mycobacteri
299	6	0.8	270	1	URED_KLEAE	Q09063	6	0.8	319	1	FIBP_ADE03	P04501	human adeno
300	6	0.8	270	1	URED_KLEPN	Q02944	6	0.8	319	1	PE13_ARATH	Q49293	arabidopsis
301	6	0.8	270	1	YNF0_YEAST	Q02944	6	0.8	319	1	RCCR_ARATH	Q81du4	arabidopsis
302	6	0.8	273	1	OSAE_BORBU	P53952	6	0.8	320	1	HEWZ_XANAC	Q8pex0	xanthomonas
303	6	0.8	274	1	COX1_CHOBI	P50668	6	0.8	320	1	MPC2_ALCEU	P17296	alcaligenes
304	6	0.8	274	1	COX1_CHOFU	P50689	6	0.8	321	1	CCSA_MAIZE	P46659	zea mays (m
305	6	0.8	274	1	COX1_CHOOC	P50670	6	0.8	321	1	CCSA_ORYSA	P12215	oryza sativ
306	6	0.8	274	1	COX1_CHORO	P50671	6	0.8	322	1	CCSA_WHEAT	P58266	triticum ae
307	6	0.8	274	1	COX3_ALLMA	P80439	6	0.8	323	1	CCSA_LOTJA	Q9bbp4	lotus japon
308	6	0.8	275	1	CTA2_ACILW	Q33950	6	0.8	323	1	CCSA_SPIOL	Q9m3j1	spinacia ol
309	6	0.8	275	1	FABI_HELPJ	Q9zmn7	6	0.8	323	1	COBD_PSEDE	P21634	pseudomonas
310	6	0.8	276	1	YB36_AQUAE	P46759	6	0.8	323	1	IN30_PEA	Q03943	pisum sativ
311	6	0.8	276	1	RSU1_HUMAN	Q87427	6	0.8	324	1	CCS1_CAEEL	P12114	caenorhabdi
312	6	0.8	277	1	RSU1_MOUSE	Q15404	6	0.8	324	1	LEGA_RAT	P38552	rattus norv
313	6	0.8	277	1	P29K_MOUSE	Q01730	6	0.8	325	1	YXDK_BACSU	P42422	bacillus su
314	6	0.8	278	1	P29K_STRGC	P42361	6	0.8	326	1	YGGK_ECOLI	P52052	escherichia
315	6	0.8	281	1	RL2_MYCCA	P10133	6	0.8	326	1	ASCD_YERPE	P37911	versinia pe
316	6	0.8	282	1	YC53_METJA	Q58650	6	0.8	328	1	CCSA_ARATH	P56770	arabidopsis
317	6	0.8	283	1	YB89_METJA	Q58589	6	0.8	328	1	YEIR_ECOLI	P33030	escherichia
318	6	0.8	284	1	KPRS_PYRAE	Q8zu24	6	0.8	329	1	RRM1_HUMAN	Q9uet6	homo sapien
319	6	0.8	284	1	RS2_MYCGE	P47316	6	0.8	330	1	HEPB_ARATH	P43273	arabidopsis
320	6	0.8	286	1	TRT2_SHEEP	P50751	6	0.8	330	1	SAPD_SALTY	P36636	salmonella
321	6	0.8	288	1	OTX2_XENLA	Q91813	6	0.8	331	1	LDHA_CHICK	P00340	gallus gall
322	6	0.8	289	1	DAPA_METJA	Q57695	6	0.8	332	1	MDHC_ARATH	P93819	arabidopsis
323	6	0.8	291	1	HSLO_BACHD	Q3kgh4	6	0.8	332	1	MDHC_BETVU	Q9sm18	beta vulgar
324	6	0.8	291	1	HSLO_BACSU	P37565	6	0.8	332	1	MDHC_MAIZE	Q08062	zea mays (m
325	6	0.8	291	1	LIP_THELA	O59952	6	0.8	332	1	MDHC_MEDSA	O48905	medicago sa

545	6	0.8	417	1	YCHO_ECOLI	p39165	eschrichia	618	1	G6P1_STRP8	Q9p2r3	streptococc
546	6	0.8	418	1	PROA_CLOAB	Q97e62	clostridium	619	1	G6P1_STRPY	Q9a1l1	streptococc
547	6	0.8	419	1	ENO_PYRAE	Q82ye7	pyrobaculum	620	1	HEMN_BRAJA	Q31381	bradyrhizob
548	6	0.8	419	1	VEGG_HUMAN	P49767	homo sapien	621	1	UTH1_YEAST	Q36135	saccharomyc
549	6	0.8	420	1	PUR2_STRPN	Q97f98	streptococc	622	1	RATM_YEAST	Q01802	saccharomyc
550	6	0.8	420	1	PUR2_STRSU	Q97f98	streptococc	623	1	PROX_MICLUE	Q50008	mycobacteri
551	6	0.8	420	1	SYM_ARCFU	O28579	archaeoglob	624	1	HOS2_YEAST	P53096	saccharomyc
552	6	0.8	421	1	PUR2_STRP3	O8k8v4	streptococc	625	1	ENGA_ANASP	Q8yzh7	anaaena sp
553	6	0.8	421	1	PUR2_STRP8	Q8p308	streptococc	626	1	SSP1_YEAST	P38789	saccharomyc
554	6	0.8	421	1	PUR2_STRPY	Q8aiy7	streptococc	627	1	SSP2_YEAST	Q12153	saccharomyc
555	6	0.8	421	1	Y118_TREPA	O83155	treponema p	628	1	TNA1_SYMTH	P31014	symbiobacte
556	6	0.8	422	1	HEAD_BPAPS	Q9t1s4	bacterioph	629	1	GLNA_METJA	Q60182	methanococc
557	6	0.8	422	1	PUR2_BACSU	P12039	bacillus su	630	1	GUAD_HUMAN	Q9yzt3	homo sapien
558	6	0.8	423	1	PUR2_NEIMA	Q9jwu6	neisseria m	631	1	PHR_STRGR	P12768	streptomyce
559	6	0.8	423	1	PUR2_NEIMB	Q9jxa3	neisseria m	632	1	SYN_MYCPN	P75521	mycoplasma
560	6	0.8	424	1	CUBO_CAEEL	O8wr51	caenorhabdi	633	1	YAO6_SCHPO	P54885	schizosacch
561	6	0.8	424	1	THIL_RAT	P17764	rattus norv	634	1	PROA_YEAST	P40085	schizosacch
562	6	0.8	425	1	FOXN_DROME	P23758	drosophila	635	1	REP_STRLI	P24805	saccharomyc
563	6	0.8	425	1	TANK_HUMAN	Q92844	homo sapien	636	1	CD4_SAISC	P22406	streptomyce
564	6	0.8	426	1	SLS1_YARLI	Q99158	yarrowia li	637	1	MURD_DEIRA	Q9rrj4	drosophila
565	6	0.8	426	1	YMW8_YEAST	Q04749	saccharomyc	638	1	DR80_DROME	O61305	schizosacch
566	6	0.8	427	1	CYB6_CHLIT	O59297	chlorobium	639	1	DVR_SCHPO	P36591	schizosacch
567	6	0.8	427	1	CYB6_CHLTE	Q9f721	chlorobium	640	1	PEX2_PICPA	Q01964	pichia past
568	6	0.8	427	1	GATD_AERPE	Q9y9t8	aeropyrum p	641	1	DALD_YERPE	P58709	versinia pe
569	6	0.8	427	1	TCO2_HUMAN	P20082	homo sapien	642	1	GALP_ECOLI	P37021	eschrichia
570	6	0.8	427	1	TCO2_RAT	Q9r086	rattus norv	643	1	Y065_MYCPN	P75612	mycoplasma
571	6	0.8	427	1	THIL_HUMAN	P24752	homo sapien	644	1	DLD3_PSEPU	P31046	pseudomonas
572	6	0.8	428	1	HEMA_CVMA5	P31615	murine coro	645	1	DNAA_PROMI	P22837	proteus mir
573	6	0.8	430	1	TCO2_MOUSE	O88968	mus musculu	646	1	Y065_MYCGE	P47311	mycoplasma
574	6	0.8	431	1	AROQ_AQUAE	O67494	aquifex aeo	647	1	TM11_MOUSE	Q99p22	mus musculu
575	6	0.8	431	1	P2X6_HUMAN	O15547	homo sapien	648	1	SG3_HUMAN	Q8wx22	homo sapien
576	6	0.8	431	1	PUR8_BACSU	P12047	bacillus su	649	1	ARI1_MOUSE	Q921k5	mus musculu
577	6	0.8	431	1	SYW_XANCP	Q8p324	xanthomonas	650	1	SECY_SULSO	Q9ux84	sulfolobus
578	6	0.8	431	1	TIG_CLOAB	Q97ft6	clostridium	651	1	YOJ8_CAEEL	P34631	caenorhabdi
579	6	0.8	431	1	UL78_HCMVA	P16751	human cytom	652	1	HQCT_RAUSE	Q9ar73	rauvolfia s
580	6	0.8	432	1	TCO2_BOVIN	O9xsc9	bos taurus	653	1	CP71_PERAE	P24465	persea amer
581	6	0.8	433	1	SYD_HALVO	O24822	halobacteri	654	1	GBA1_YEAST	P08539	saccharomyc
582	6	0.8	433	1	AKR_ARATH	Q05753	arabidopsis	655	1	FXH1_BRARE	Q919el	brachydanio
583	6	0.8	435	1	DHOM_METGL	P37143	methylobaci	656	1	GLGS_WHEAT	P30523	tritium ae
584	6	0.8	435	1	FUZ7_USTWA	Q99078	ustillago ma	657	1	MAIR_YEAST	P53338	saccharomyc
585	6	0.8	436	1	SH6_RAT	P31388	rattus norv	658	1	MA6R_YEAST	P10508	saccharomyc
586	6	0.8	436	1	ENV_FLVCS	Q02077	feline leuk	659	1	CREC_ECOLI	P08401	eschrichia
587	6	0.8	436	1	BRFA_ARATH	Q39097	arabidopsis	660	1	GAB2_HUMAN	P47870	homo sapien
588	6	0.8	437	1	AP2A_HUMAN	P05549	homo sapien	661	1	GAB2_MOUSE	P15432	mus musculu
589	6	0.8	437	1	AP2A_MOUSE	P34056	mus musculu	662	1	KLF4_MOUSE	Q60793	mus musculu
590	6	0.8	437	1	AP2A_RAT	P58197	rattus norv	663	1	PTGH_ECOLI	P05053	eschrichia
591	6	0.8	437	1	ERF1_HUMAN	P46055	homo sapien	664	1	PYR5_PICDI	P09556	dictyosteli
592	6	0.8	437	1	ERF1_XENLA	P35615	xenopus lae	665	1	S61A_SCHPO	P79088	schizosacch
593	6	0.8	437	1	SECY_STRCO	P46785	streptomyce	666	1	PTMB_CLOAB	O65989	clostridium
594	6	0.8	437	1	SECY_STRGB	O59912	streptomyce	667	1	DUSA_HUMAN	Q9Y5W6	homo sapien
595	6	0.8	437	1	SECY_STRGR	O59916	streptomyce	668	1	DUSA_MOUSE	Q9esa0	mus musculu
596	6	0.8	437	1	SECY_STRLI	P49977	streptomyce	669	1	GLGS_ORISA	P15280	oryza sativ
597	6	0.8	437	1	SECY_STRSC	P43416	streptomyce	670	1	SYE_DEIRA	Q9rx30	deinococcus
598	6	0.8	438	1	GATD_PYRAB	Q9V0E9	pyrococcus	671	1	NCAP_HAZVU	P27318	hazara viru
599	6	0.8	438	1	GATD_PYRHO	O59132	pyrococcus	672	1	BCHZ_RHOGE	Q97jbb	rhodocyclu
600	6	0.8	439	1	EX7L_HAEIN	P43913	haemophilus	673	1	MDM2_CANFA	P56950	canis fami
601	6	0.8	440	1	SH6_HUMAN	P50406	homo sapien	674	1	GALT_LACHE	Q00054	lactobacill
602	6	0.8	440	1	SH6_MOUSE	Q9rlc8	mus musculu	675	1	YKT1_CAEEL	P06054	caenorhabdi
603	6	0.8	440	1	MURD_VIECH	Q8kpgs	vibrio chol	676	1	DHAB_ECOLI	P34312	caenorhabdi
604	6	0.8	441	1	TONE_COCCA	O74205	cochliobolu	677	1	GLGS_BETVU	P17445	eschrichia
605	6	0.8	443	1	GLNA_CLOSA	P10656	clostridium	678	1	ET83_DROME	P55232	beta vulgar
606	6	0.8	444	1	GUNN_ERWCA	Q59394	erwinia car	679	1	AP2A_SHEEP	P29774	drosophila
607	6	0.8	445	1	DMSA_ACHLA	Q9khu8	acholeplasm	680	1	TYTR_TYCR	Q9n0n3	ovis aries
608	6	0.8	445	1	MRSA_HAEIN	P45164	haemophilus	681	1	ANL2_MOUSE	P28593	trypanosoma
609	6	0.8	446	1	GLNA_METVO	P21154	methanococc	682	1	ANYA_DROMA	Q9r045	mus musculu
610	6	0.8	446	1	PIV2_ADECC	Q85945	canine aden	683	1	ANYA_DROME	P54215	drosophila
611	6	0.8	446	1	PIV2_ADECT	P87552	canine aden	684	1	ANYB_DROME	P08144	drosophila
612	6	0.8	447	1	CLUS_RAT	Q05371	rattus norv	685	1	TEBA_OXYNO	P81641	drosophila
613	6	0.8	447	1	DTA3_RALSO	Q8xr66	ralstonia s	686	1	LYXK_ECOLI	P29549	oxytricha n
614	6	0.8	447	1	HWC8_SCHPO	P54874	schizosacch	687	1	OPSI_PATYE	P37677	eschrichia n
615	6	0.8	448	1	CLUS_MOUSE	O65890	mus musculu	688	1	DP22_RAT	Q15973	patinopecte
616	6	0.8	448	1	HEM1_METWA	Q8pw60	methanosarc	689	1	SECD_AQUAE	Q67102	aquifex aeo
617	6	0.8	449	1	G6P1_STRF3	Q8k8q6	streptococc	690	1	PIV2_ADECR	Q96680	canine aden

691	6	0.8	503	1	CATA_MICLU	P29422 micrococcus	764	0.8	566	1	CC45_HUMAN	O75419 homo sapien
692	6	0.8	503	1	SECD_HELPY	Q26074 helicobacte	765	0.8	566	1	CC45_MOUSE	Q261x9 mus musculus
693	6	0.8	503	1	SNAL_MOUSE	Q61234 mus musculus	766	0.8	569	1	ATKB_SALTI	Q282e5 salmonella
694	6	0.8	504	1	GUNW_ERWCA	Q59395 erwinia car	767	0.8	569	1	HYCE_ECOLI	P16431 escherichia
695	6	0.8	505	1	GUNW_ERWCA	Q47096 erwinia car	768	0.8	572	1	GPCS_HUMAN	P78333 homo sapien
696	6	0.8	505	1	SNAL_HUMAN	Q13424 homo sapien	769	0.8	575	1	ESR2_ICTPU	Q91aki ictalurus p
697	6	0.8	505	1	SNAL_RABIT	Q28626 oryctolagus	770	0.8	576	1	ACEA_RICCO	P15479 ricinus com
698	6	0.8	505	1	YOD2_SCHPO	Q9c0v0 schizosacch	771	0.8	576	1	ACH2_DROME	P17644 drosophila
699	6	0.8	506	1	DLTA_LACRH	P71338 haemophilus	772	0.8	576	1	GRK6_HUMAN	P43250 homo sapien
700	6	0.8	506	1	HITB_HABIN	P71338 haemophilus	773	0.8	576	1	GRK6_MOUSE	O70293 mus musculus
701	6	0.8	507	1	CHLB_PORPU	P51278 porphyra pu	774	0.8	576	1	GRK6_RAT	P97711 rattus norv
702	6	0.8	508	1	CC37_CANAL	Q81266 candida alb	775	0.8	576	1	RMIN_DROME	P97711 rattus norv
703	6	0.8	508	1	GLGS_VICFA	P52416 vicia faba	776	0.8	577	1	SYR_SALTI	P05527 drosophila
704	6	0.8	509	1	C931_SOYEN	Q42798 glycine max	777	0.8	577	1	SYR_SALTY	Q825v7 salmonella
705	6	0.8	510	1	C933_SOYEN	O81973 glycine max	778	0.8	577	1	YGSU_YEAST	P74871 salmonella
706	6	0.8	511	1	EGO_ECOLI	P77257 escherichia	779	0.8	577	1	YCSJ_BACSU	P53333 saccharomyc
707	6	0.8	512	1	G63A_DROME	Q9vz17 drosophila	780	0.8	578	1	PHIA_ECOLI	P42967 bacillus su
708	6	0.8	512	1	GLGT_VICFA	P52417 vicia faba	781	0.8	579	1	RMDH_MAIZE	Q47153 escherichia
709	6	0.8	513	1	GLGS_HORVU	P25238 hordeum vul	782	0.8	580	1	ACEA_PINTA	O24594 zea mays (m
710	6	0.8	513	1	SYS_METHH	P527194 methanobact	783	0.8	584	1	ACES_RABIT	Q29499 oryctolagus
711	6	0.8	514	1	CBH_SCHPO	O14423 schizosacch	784	0.8	586	1	CO9_FUGRU	P79755 fugu rubrip
712	6	0.8	514	1	CPT7_ICTPU	Q73853 ictalurus p	785	0.8	588	1	CALI_BOVIN	Q28068 bos taurus
713	6	0.8	514	1	IMA3_CABEL	Q19969 caenorhabdi	786	0.8	588	1	CALI_HUMAN	Q13939 homo sapien
714	6	0.8	515	1	LEU1_HABIN	P43861 haemophilus	787	0.8	588	1	DECA_DROME	P07713 drosophila
715	6	0.8	515	1	YWCA_BACSU	P29599 bacillus su	788	0.8	590	1	GRK5_HUMAN	P34947 homo sapien
716	6	0.8	516	1	LADI_HUMAN	O00515 homo sapien	789	0.8	590	1	GRK5_RAT	Q62833 rattus norv
717	6	0.8	518	1	LEU1_PASMO	Q9cjm5 pasteurella	790	0.8	590	1	SYD_SALTY	Q825w1 salmonella
718	6	0.8	518	1	SAP_CHICK	O13035 gallus gall	791	0.8	590	1	SYD_PSEAR	Q82mv2 salmonella
719	6	0.8	518	1	TEX5_HUMAN	Q99593 homo sapien	792	0.8	591	1	SYD_VIECH	Q51422 pseudomonas
720	6	0.8	518	1	TEX5_MOUSE	P70326 mus musculus	793	0.8	591	1	SYD_VIBU	Q87q2 vibrio chol
721	6	0.8	520	1	CZCB_ALCEU	P13510 alcaligenes	794	0.8	592	1	SYD_VIBPA	Q87q2 vibrio chol
722	6	0.8	520	1	CZCB_ALCEU	P41176 alcaligenes	795	0.8	594	1	NU5M_HIPAM	Q92zy1 hippopotamu
723	6	0.8	520	1	GLGS_ARATH	P52228 arabidopsis	796	0.8	595	1	BETP_CORGL	P54582 corynebacte
724	6	0.8	520	1	GLGS_BRANA	Q9m462 brassica na	797	0.8	595	1	BPRX_BACNO	P42780 bacteroides
725	6	0.8	521	1	GLGS_LYCES	Q42882 lycopersico	798	0.8	595	1	DCOA_KLEPN	P13187 klebsiella
726	6	0.8	521	1	GLGS_SOLTU	P23509 solanum tub	799	0.8	600	1	ESRI_RAT	P06211 rattus norv
727	6	0.8	525	1	GOR_PANTR	P48778 pan troglod	800	0.8	602	1	YZ13_METJA	Q60275 methanococc
728	6	0.8	526	1	SECD_HELPY	Q9zj66 helicobacte	801	0.8	603	1	PURL_THEMA	Q9x0x3 thermotoga
729	6	0.8	527	1	NPRE_BACER	P43263 bacillus br	802	0.8	604	1	YJ13_YEAST	P47030 saccharomyc
730	6	0.8	529	1	YRYP_GIALA	P25203 giardia lam	803	0.8	605	1	PESC_YEAST	P53261 saccharomyc
731	6	0.8	532	1	CATA_DEIRA	Q9u915 drosophila	804	0.8	606	1	CACI_YEAST	Q12495 saccharomyc
732	6	0.8	532	1	CRTI_AHPSP	P21134 aphanocepca	805	0.8	607	1	HEMA_PHODV	P28862 phocine dis
733	6	0.8	537	1	CEKI_HUMAN	Q59337 deinococcus	806	0.8	609	1	XYNA_NEOPA	Q11146 bacterioph
734	6	0.8	537	1	CK11_YEAST	Q8ctc0 homo sapien	807	0.8	609	1	YJ15_MYCLE	P46509 mycobacteri
735	6	0.8	538	1	NPRB_BACSU	P23291 saccharomyc	808	0.8	610	1	POUC_BRARE	P21367 brachydanio
736	6	0.8	538	1	NPRB_BACSU	P39899 bacillus su	809	0.8	611	1	ACES_FELCA	O62763 felis silve
737	6	0.8	540	1	G6PI_CORGL	Q9ns31 corynebacte	810	0.8	611	1	PHBC_RHIME	P50176 r poly-beta
738	6	0.8	540	1	GTRP_HUMAN	Q9nm0 homo sapien	811	0.8	613	1	ACES_BOVIN	P22303 homo sapien
739	6	0.8	541	1	60IM_HABIN	P44973 haemophilus	812	0.8	614	1	ACES_MOUSE	P21836 mus musculu
740	6	0.8	542	1	DAGA_ALTHA	P30144 alteromonas	813	0.8	614	1	ACES_RAT	P37136 rattus norv
741	6	0.8	543	1	CHK2_HUMAN	Q96017 homo sapien	814	0.8	615	1	DNAX_THETH	Q56235 thermus the
742	6	0.8	543	1	DALS_YEAST	P15365 saccharomyc	815	0.8	615	1	MUTA_MCTTU	P71773 mycobacteri
743	6	0.8	543	1	TIMD_MYCPN	P75436 mycoplasma	816	0.8	616	1	ACOC_SOLTU	O04916 solanum tub
744	6	0.8	544	1	ALG6_YEAST	Q12001 saccharomyc	817	0.8	616	1	RRPO_PLRV1	P17520 potato leaf
745	6	0.8	545	1	CH60_SODGL	Q9anr8 sodalis glo	818	0.8	616	1	Y396_HELPY	Q92kf3 helicobacte
746	6	0.8	547	1	CATX_BACSU	P94377 bacillus su	819	0.8	616	1	GIDA_CAUCR	O25157 helicobacte
747	6	0.8	548	1	VG31_BPMU5	O05240 mycobacteri	820	0.8	618	1	ORCC_DROME	Q9xbf8 caulobacter
748	6	0.8	550	1	COCH_HUMAN	Q43405 homo sapien	821	0.8	618	1	GLMS_CYACA	Q24168 drosophila
749	6	0.8	550	1	MANB_MYCGE	P47299 mycoplasma	822	0.8	620	1	IORA_ARCFU	O19908 c glucosami
750	6	0.8	551	1	ASLA_ECOLI	P25549 escherichia	823	0.8	622	1	VP27_YEAST	P40343 saccharomyc
751	6	0.8	551	1	Y275_HABIN	P43975 haemophilus	824	0.8	623	1	PTN1_DROME	O28783 archaeoglob
752	6	0.8	552	1	AP1_SCHPO	Q01663 schizosacch	825	0.8	623	1	RECG_HELPY	P51022 drosophila
753	6	0.8	552	1	COCH_MOUSE	Q62507 mus musculu	826	0.8	623	1	RECG_HELPY	Q9zjai helicobacte
754	6	0.8	552	1	VIE2_HSVB4	Q02479 bovine herp	827	0.8	624	1	STS_MOUSE	O26051 helicobacte
755	6	0.8	555	1	HYFG_ECOLI	P77329 escherichia	828	0.8	626	1	GLMS_NOSS9	P50427 mus musculu
756	6	0.8	557	1	ACHI_SCHGR	P23414 schistocerc	829	0.8	628	1	HNFA_MOUSE	O68280 n glucosami
757	6	0.8	557	1	AR11_HUMAN	Q9y4x5 homo sapien	830	0.8	629	1	RA21_XENLA	P15257 rattus norv
758	6	0.8	557	1	SCP2_MOUSE	Q9cun3 mus musculu	831	0.8	629	1	SYM_THEMEA	O33925 thermotoga
759	6	0.8	563	1	ARAB_MYCSM	Q9lbd3 mycobacteri	832	0.8				
760	6	0.8	564	1	FRP1_SCHPO	Q04800 schizosacch	833	0.8				
761	6	0.8	564	1	OXC_ECOLI	P78093 escherichia	834	0.8				
762	6	0.8	564	1	SYT_MYCGE	P47615 mycoplasma	835	0.8				
763	6	0.8	565	1	Y912_ARCFU	O29350 archaeoglob	836	0.8				

983 6 0.8 781 1 TL22_CHICK
 984 6 0.8 782 1 GELS_HUMAN
 985 6 0.8 785 1 UBPI_HUMAN
 986 6 0.8 790 1 KIF9_HUMAN
 987 6 0.8 790 1 SYFB_CHIMU
 988 6 0.8 790 1 TNP3_HUMAN
 989 6 0.8 793 1 TL21_CHICK
 990 6 0.8 797 1 D151_HAIRIN
 991 6 0.8 800 1 PT27_YEAST
 992 6 0.8 802 1 OPH1_HUMAN
 993 6 0.8 802 1 OPH1_MOUSE
 994 6 0.8 802 1 PF8C_SALTY
 995 6 0.8 802 1 PFN8_MOUSE
 996 6 0.8 802 1 S15A_MOUSE
 997 6 0.8 803 1 ATFCU_BAGSU
 998 6 0.8 803 1 SYFB_WIGBR
 999 6 0.8 804 1 S15A_HUMAN
 1000 6 0.8 804 1 S15A_RAT

ALIGNMENTS

RESULT 1

PE42_ARATH STANDARD; PRT; 330 AA.
 AC Q95B81; Q41937; Q42579; Q43730;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Peroxidase 42 precursor (EC 1.11.1.7) (Atperox P42) (PRXR1)
 DE (ATP1a/ATP1b).
 GN PER42 OR P42 OR AT4G21960 OR T805.170 OR FIN20.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX STRAIN=cv. Columbia; TISSUE=Flower, Leaf, Root, Siliques, and Stem;
 RA Capelli N., Tognolli M., Flach J., Overney S., Penel C., Greppin H.,
 RA Simon P.;
 RA "Eleven cDNA clones from Arabidopsis thaliana encoding
 RT isoperoxidases";
 RL (In) Plant Gene Register FGR96-066.
 RN [2]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX MEDLINE=97239476; PubMed=9132061;
 RA Kjaergaard I.V.H., Jespersen H.M., Rasmussen S.K., Welinder K.G.;
 RT "Sequence and RT-PCR expression analysis of two peroxidases from
 RT Arabidopsis thaliana belonging to a novel evolutionary branch of plant
 RT peroxidases";
 RL Plant Mol. Biol. 33:699-708(1997).
 RN [3]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoerg W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
 RA Kreis M., Delsen M., Fuldgonen P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Billam L., Robben J.,
 RA Van der Schueren J., Grynoprez B., Chuang Y.-J., Vandenbussche P.,
 RA Braeken M., Weljens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,

De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 Clark L., Doggett J., Hall S., Kay M., Lennard N., McLaughlin K., Mayes R.,
 Pettett A., Rajandream M.A., Lyne M., Benes V., Rehnemann S.,
 Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
 Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
 Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 Schnabl S., Hiller R., Schmidt W., Lecharry A., Aubourg S.,
 Chedof S., Cooke R., Berger C., Monfort A., Casacuberta E.,
 Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 Heijman D., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,
 Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., L.,
 Kramer J., Fulten L., Mardis E., Dante M., Pepin K., Hillier L.,
 Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 Chen B., Marra M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana";
 RL Nature 402:769-777(1999).
 RN [4]
 RC SEQUENCE FROM N.A.
 RP Chao Q., Brooks S.Y., Chen H., Johnson-Hopson C., Khan S., Shinn P.,
 RA Ecker J.R.;
 RT "Full length cDNA sequence of Arabidopsis thaliana";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PSEC)";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RC SEQUENCE OF 1-168 FROM N.A.
 RP TISSUE=Leaf;
 RA Stracke R., Palme K.;
 RT "Signal peptide selection derived cDNAs from Arabidopsis thaliana
 RT leaves and guard cells";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RC SEQUENCE OF 278-330 FROM N.A.
 RP STRAIN=cv. Columbia; TISSUE=Green siliques;
 RA Raynal M., Greillet F., Laude M., Meyer Y., Cooke R., Delsen M.;
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RC CHARACTERIZATION
 RP STRAIN=cv. Columbia;
 RX MEDLINE=98409321; PubMed=9738941;
 RA Oestergaard L., Pedersen A.G., Jespersen H.M., Brunak S.,
 RA Welinder K.G.;
 RT "Computational analyses and annotations of the Arabidopsis peroxidase
 RT gene family";
 RL FEBS Lett. 433:98-102(1998).
 RN [9]
 RC TISSUE SPECIFICITY
 RP STRAIN=cv. Columbia;
 RX Zhu T., Budworth P., Han B., Brown D., Chang H.-S., Zou G., Wang X.;
 RT "Toward elucidating the global gene expression patterns of developing
 RT Arabidopsis: parallel analysis of 8300 genes by a high-density
 RT oligonucleotide probe array";

RL Plant Physiol. Biochem. 39:221-242 (2001).
 RP [10]
 RP INDUCTION.
 RC STRAIN=cv. Columbia;
 RX PubMed=11158533;
 RA Schaffer R., Landgraf J., Accerbi M., Simon V., Larson M., Wisman E.;
 RT "Microarray analysis of diurnal and circadian-regulated genes in
 RT Arabidopsis";
 RL Plant Cell 13:113-123 (2001).
 RN [11]
 RN GENE FAMILY ORGANIZATION, AND NOMENCLATURE.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=22030461; PubMed=12034502;
 RA Tognolli M., Penel C., Greppin H., Simon P.;
 RT "Analysis and expression of the class III peroxidase large gene family
 RT in Arabidopsis thaliana";
 RL Gene 288:129-138 (2002).
 CC -/- FUNCTION: Removal of H(2)O(2), oxidation of toxic reductants,
 CC biosynthesis and degradation of lignin, suberization, auxin
 CC catabolism, response to environmental stresses such as wounding,
 CC pathogen attack and oxidative stress. These functions might be
 CC dependent on each isozyme/isoform in each plant tissue.
 CC -/- FUNCTION: Might function as heat shock-like defense protein.
 CC -/- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
 CC -/- COFACTOR: Binds 1 protoheme IX, 1 iron(III) ion and 2 calcium
 CC ions
 CC -/- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -/- TISSUE SPECIFICITY: Constitutively expressed in the whole plant,
 CC with the highest expression in roots.
 CC -/- INDUCTION: Expressed under a diurnal rhythm (circadian clock
 CC control).
 CC -/- MISCELLANEOUS: There are 73 peroxidase genes in A.thaliana.
 CC -/- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. CLASSICAL PLANT
 CC (CLASS III) PEROXIDASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 DR ENBL; X98113; CAA66957.1; -;
 DR ENBL; X98189; CAA68862.1; -;
 DR ENBL; AL021890; CAA17163.1; ALT_INIT.
 DR ENBL; AL161556; CAB79151.1; ALT_INIT.
 DR ENBL; AL022140; -; NOT_ANNOTATED_CDS.
 DR ENBL; AF325015; AAC40367.1; -;
 DR ENBL; AF428379; AAL16147.1; -;
 DR ENBL; AY056809; AAL10500.1; -;
 DR ENBL; AY058071; AAL24179.1; -;
 DR ENBL; AY059810; AAL24292.1; -;
 DR ENBL; AY132009; AAM91042.1; -;
 DR ENBL; AF083767; AAN60325.1; -;
 DR ENBL; Z17792; CAA79071.1; -;
 DR HSP; P00433; 2ATJ
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF00141; peroxidase; 1.
 DR PRINTS; PR00458; PEROXIDASE.
 DR PROSITE; PS00435; PEROXIDASE_1; FALSE_NEG.
 DR PROSITE; PS00436; PEROXIDASE_2; FALSE_NEG.
 DR PROSITE; PS50873; PEROXIDASE_4; 1.
 KW Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;
 KW Multigene family; Calcium; Signal; Biological rhythms.
 FT SIGNAL 1 23
 FT CHAIN 24 330 PEROXIDASE 42.
 FT METAL 72 72 CALCIUM 1 (BY SIMILARITY).
 FT METAL 75 75 CALCIUM 1 (VIA CARBONYL OXYGEN)
 FT (BY SIMILARITY).
 FT METAL 79 79 CALCIUM 1 (BY SIMILARITY).
 FT METAL 81 81 CALCIUM 1 (BY SIMILARITY).
 FT METAL 198 198 CALCIUM 2 (BY SIMILARITY).

FT METAL 247 247 CALCIUM 2 (BY SIMILARITY).
 FT METAL 250 250 CALCIUM 2 (BY SIMILARITY).
 FT METAL 255 255 CALCIUM 2 (BY SIMILARITY).
 FT ACT_SITE 67 67 BY SIMILARITY.
 FT ACT_SITE 71 71 DISTAL HISTIDINE.
 FT ACT_SITE 167 167 SUBSTRATE BINDING (BY SIMILARITY).
 Query Match 1.2%; Score 9; DB 1; Length 330;
 Best Local Similarity 100.0%; Pred. No. 0.37;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 408 ALSATSEAV 416
 DB 16 ALSATSEAV 24
 RESULT 2
 MY9B RAT STANDARD; PRT; 1980 AA.
 AC Q63358;
 LR 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Myosin IXb (Unconventional myosin-9b).
 GN MYO9B OR MYR5.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain stem, and Spinal cord;
 RX MEDLINE=95188874; PubMed=7882973;
 RA Reinhard J., Scheel A.A., Diekmann D., Hall A., Ruppert C.,
 RA Baehler M.;
 RT "A novel type of myosin implicated in signalling by rho family
 RT GTPases";
 RL EMBO J. 14:697-704 (1995).
 CC -/- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
 CC MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS
 CC ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP
 CC AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
 CC ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.
 CC -/- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH
 CC F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS
 CC LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE
 CC PERINUCLEAR REGION (BY SIMILARITY).
 CC -/- TISSUE SPECIFICITY: EXPRESSED IN TESTIS, LUNG, THYMUS, BRAIN,
 CC LIVER, AND SPLEEN.
 CC -/- SIMILARITY: Contains 4 IQ domains.
 CC -/- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -/- SIMILARITY: Contains 1 Ras-associating domain.
 CC -/- SIMILARITY: Contains 1 Rho-GAP domain.
 CC -/- SIMILARITY: Contains 1 zinc-dependent phorbol-ester and DAG
 CC binding domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 DR ENBL; X77609; CAA54700.1; -;
 DR PIR; S54307; S54307.
 DR HSP; P08799; IAMD.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin head.
 DR InterPro; IPR000159; RA domain.
 DR InterPro; IPR000198; RhoGAP.

DR Pfam; PF00130; DAG_PE-bind; 1.
 DR Pfam; PF00612; IQ; 4.
 DR Pfam; PF00063; myosin_head; 2.
 DR Pfam; PF00788; RA; 1.
 DR Pfam; PF00620; RhoGAP; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRODOM; PD000355; myosin_head; 1.
 DR SMART; SM00109; C1; 1.
 DR SMART; SM00015; IQ; 3.
 DR SMART; SM00242; MYSC; 1.
 DR SMART; SM00314; RA; 1.
 DR SMART; SM00324; RhoGAP; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 1.
 DR PROSITE; PS50096; IQ; 3.
 DR PROSITE; PS50200; RA; 1.
 DR PROSITE; PS50238; RHO GAP; 1.
 DR Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
 KW Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;
 KW Zinc.

FT DOMAIN 1 940 HEAD OR MOTOR DOMAIN.
 FT DOMAIN 941 1045 NECK OR REGULATORY DOMAIN.
 FT DOMAIN 1046 1380 TAIL.
 FT DOMAIN 1841 1861 COILED COIL (POTENTIAL).
 FT DOMAIN 1918 1948 COILED COIL (POTENTIAL).
 FT DOMAIN 15 114 RAS-ASSOCIATING.
 FT DOMAIN 845 856 ACTIN-BINDING.
 FT DOMAIN 958 978 IQ 1.
 FT DOMAIN 981 1001 IQ 2.
 FT DOMAIN 1002 1024 IQ 3.
 FT DOMAIN 1025 1054 IQ 4.
 FT DOMAIN 1593 1641 PHORBOL-ESTER AND DAG BINDING.
 FT DOMAIN 1663 1848 RHO-GAP.
 FT NP_BIND 239 246 ATP (POTENTIAL).
 SQ SEQUENCE 1980 AA; 225035 MW; D79FEC4D0FAE0C05 CRC64;

Query Match 1.2%; Score 9; DB 1; Length 1980;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 TEGLYRKSQ 213
 Db 1690 TEGLYRKSQ 1698
 |||||

RESULT 3
 ID MY9B MOUSE STANDARD; PRT; 2114 AA.
 AC Q9QY06; Q9QY07; Q9QY08; Q9QY09;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Myosin IxB (Unconventional myosin-9b).
 GN MY9B OR MYR5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANTS.
 RC STRAIN=C57BL/6;
 RX MEDLINE=20047919; PubMed=10580159;
 RA Grewal P.K.; Jones A.-M.; Maconochie M.; Lemmers R.J.F.; Frants R.R.;
 RA Hewitt J.E.;
 RT "Cloning of the murine unconventional myosin gene Myo9b and
 RT identification of alternative splicing."
 RL Gene 240:389-398(1999).
 CC -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
 CC MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS
 CC ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP
 CC AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
 CC ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.

CC -!- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH
 CC F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS
 CC LOCALIZATION IS CYTOSOLIC WITH THE HIGHEST LEVELS IN THE
 CC PERINUCLEAR REGION (BY SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=Q9QY06-1; Sequence=Displayed;
 CC Name=2; Synonyms=Q;
 CC IsoId=Q9QY06-2; Sequence=VSP_003363;
 CC Name=3; Synonyms=C;
 CC IsoId=Q9QY06-3; Sequence=VSP_003364; VSP_003365;
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE BRAIN, HEART, MUSCLE AND
 CC INNER EAR.
 CC -!- SIMILARITY: Contains 4 IQ domains.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 Ras-associating domain.
 CC -!- SIMILARITY: Contains 1 Rho-GAP domain.
 CC -!- SIMILARITY: Contains 1 zinc-dependent phorbol-ester and DAG
 CC binding domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF143687; AAF00122.1; -;
 CC EMBL; AF143685; AAF00120.1; -;
 CC EMBL; AF143686; AAF00121.1; ALT_FRAME.
 CC EMBL; AF143683; AAF00118.1; -;
 CC HSP; P08799; 1MND.
 CC MGD; MGI:106624; Myo9b.
 CC InterPro; IPR002219; DAG_PE-bind.
 CC InterPro; IPR000048; IQ_region.
 CC InterPro; IPR001609; myosin_head.
 CC InterPro; IPR000159; RA domain.
 CC InterPro; IPR000198; RhoGAP.
 CC Pfam; PF00130; DAG_PE-bind; 1.
 CC Pfam; PF00612; IQ; 4.
 CC Pfam; PF00063; myosin_head; 2.
 CC Pfam; PF00788; RA; 1.
 CC Pfam; PF00620; RhoGAP; 1.
 CC PRINTS; PR00193; MYOSINHEAVY.
 CC PRODOM; PD000355; myosin_head; 1.
 CC SMART; SM00109; C1; 1.
 CC SMART; SM00015; IQ; 4.
 CC SMART; SM00242; MYSC; 1.
 CC SMART; SM00314; RA; 1.
 CC SMART; SM00324; RhoGAP; 1.
 CC PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
 CC PROSITE; PS00081; DAG_PE_BIND_DOM_2; 1.
 CC PROSITE; PS50096; IQ; 3.
 CC PROSITE; PS50200; RA; 1.
 CC PROSITE; PS50238; RHO GAP; 1.
 CC Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
 KW Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;
 KW Zinc; Alternative splicing; Polymorphism.
 FT DOMAIN 1 939 HEAD OR MOTOR DOMAIN.
 FT DOMAIN 940 1044 NECK OR REGULATORY DOMAIN.
 FT DOMAIN 1045 2114 TAIL.
 FT DOMAIN 1562 1592 COILED COIL (POTENTIAL).
 FT DOMAIN 1839 1859 COILED COIL (POTENTIAL).
 FT DOMAIN 1915 1945 COILED COIL (POTENTIAL).
 FT DOMAIN 15 114 RAS-ASSOCIATING.
 FT DOMAIN 844 855 ACTIN-BINDING.
 FT DOMAIN 957 977 IQ 1.
 FT DOMAIN 979 1000 IQ 2.
 FT DOMAIN 1001 1023 IQ 3.
 FT DOMAIN 1024 1053 IQ 4.
 FT DOMAIN 1591 1639 PHORBOL-ESTER AND DAG BINDING.

CC	-1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS. MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
CC	-1- ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.
CC	-1- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE PERINUCLEAR REGION.
CC	-1- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Name=Long;
CC	isoId=Q13459-1; Sequence=Displayed;
CC	Name=Short;
CC	isoId=Q13459-2; Sequence=VSP_003361, VSP_003362;
CC	-1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN PERIPHERAL BLOOD LEUKOCYTES AND AT LOWER LEVELS, IN THYMUS, SPLEEN, TESTIS, PROSTATE, OVARY, BRAIN, SMALL INTESTINE, AND LUNG.
CC	-1- SIMILARITY: Contains 4 IQ domains.
CC	-1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC	-1- SIMILARITY: Contains 1 Ras-associating domain.
CC	-1- SIMILARITY: Contains 1 Rho-GAP domain.
CC	-1- SIMILARITY: Contains 1 zinc-dependent phorbol-ester and DAG binding domain.
CC	-1- CAUTION: THE C-TERMINAL SEQUENCE FROM AA 1917 ONWARDS FROM REF.1 WAS PROBABLY A CHIMERA.
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	EMBL; U42391; AAC50402.1; ALT_SEQ.
DR	EMBL; AF143684; AAF00119.1; -.
DR	EMBL; AF020267; AAC26597.1; -.
DR	EMBL; AK002201; BAA92132.1; ALT_INIT.
DR	HSSP; P08799; IMND.
DR	Genew; HGNC:7609; MYO9B.
DR	MIM; 602129; -.
DR	GO; GO:0005653; C:perinuclear space; TAS.
DR	GO; GO:0003774; F:motor activity; TAS.
DR	GO; GO:0008570; F:myosin ATPase activity; TAS.
DR	GO; GO:0005100; F:rho GTPase activator activity; TAS.
DR	InterPro; IPR002219; DAG_PE-bind.
DR	InterPro; IPR000048; IQ region.
DR	InterPro; IPR001509; myosin head.
DR	InterPro; IPR000198; RhoGAP.
DR	Pfam; PF00130; DAG_PE-bind; 1.
DR	Pfam; PF00612; IQ; 4.
DR	Pfam; PF00063; myosin_head; 2.
DR	Pfam; PF00788; RA; 1.
DR	Pfam; PF00620; RhoGAP; 1.
DR	PRINTS; PR00193; MYOSINHEAVY.
DR	ProDom; PD000355; myosin_head; 1.
DR	SMART; SM00109; C1; 1.
DR	SMART; SM00015; IQ; 4.
DR	SMART; SM00242; MYSC; 1.
DR	SMART; SM00314; RA; 1.
DR	SMART; SM00324; RhoGAP; 1.
DR	PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR	PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR	PROSITE; PS50096; IQ; 3.
DR	PROSITE; PS50200; RA; 1.
DR	PROSITE; PS50238; RHO GAP; 1.
KW	Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW	Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;
KW	Zinc; Alternative splicing.
FT	DOMAIN 1 939 HEAD OR MOTOR DOMAIN.

FT DOMAIN 940 1044 NECK OR REGULATORY DOMAIN.
FT DOMAIN 1045 2158 TAIL.
FT DOMAIN 1046 1071 COILED COIL (POTENTIAL).
FT DOMAIN 1071 1880 COILED COIL (POTENTIAL).
FT DOMAIN 1880 1901 COILED COIL (POTENTIAL).
FT DOMAIN 1959 1989 COILED COIL (POTENTIAL).
FT DOMAIN 15 114 RAS-ASSOCIATING.
FT DOMAIN 844 855 ACTIN-BINDING.
FT DOMAIN 957 977 IQ 1.
FT DOMAIN 979 1000 IQ 2.
FT DOMAIN 1001 1023 IQ 3.
FT DOMAIN 1024 1053 IQ 4.
FT DOMAIN 1633 1681 PHOBOL-ESTER AND DAG BINDING.
FT DOMAIN 1703 1888 RHOR-GAP.
FT NP BIND 239 246 ATP (POTENTIAL).
FT VARSPLIC 2022 2023 /FTID=VSP_003361.
FT VARSPLIC 2024 2158 Missing (in isoform Short).
FT CONFLICT 1937 1939 QVP -> KT (IN REF. 4).
FT CONFLICT 1947 1947 L -> P (IN REF. 4).
FT CONFLICT 2040 2045 TVAAP -> PWPELH (IN REF. 3).
FT CONFLICT 2049 2049 P -> L (IN REF. 3).
FT CONFLICT 2067 2067 P -> S (IN REF. 3).
FT CONFLICT 2157 2158 NG -> MAESH (IN REF. 3).
SQ SEQUENCE 2158 AA; 243556 MW; 4978FID770F56D28 CRC64;

Query Match 1.2%; Score 9; DB 1; Length 2158;
Best Local Similarity 100.0%; Pred.No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TEGLYRKSQ 213
Db 1730 TEGLYRKSQ 1738
|||||

RESULT 5
Y096 AGRT5
ID Y096 AGRT5 STANDARD; PRT; 196 AA.
AC P58792;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative NADH dehydrogenase/NAD(P)H nitroreductase Atu2496
(EC 1.1.1.17).
DE (EC 1.1.1.17).
GN ATU2496 OR AGR_C_4536.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavon T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liac L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58";
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,

RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58";
RL Science 294:2323-2328 (2001).
CC -1- COFACTOR: FMN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE NITROREDUCTASE FAMILY. HADB SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE009196; AAL43483.1; -
DR EMBL; AE008162; AAK88228.1; -
DR PIR; AE2883; AE2883.
DR PIR; C97659; C97659.
DR HAMAP; MF 01204; -; 1.
DR InterPro; IPR000415; Nitroreductase.
DR Pfam; PF00881; Nitroreductase; 1.
KW Hypothetical protein; Oxidoreductase; NAD; NADP; Flavoprotein; FMN;
KW Complete proteome.
SQ SEQUENCE 196 AA; 21139 MW; 6283E17567796984 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 196;
Best Local Similarity 100.0%; Pred.No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 LSSGNLEK 187
Db 70 LSSGNLEK 77
|||||

RESULT 6
BLA3_BACCE
ID BLA3_BACCE STANDARD; PRT; 316 AA.
AC P06548;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Beta-lactamase III precursor (EC 3.5.2.6).
GN BLA3.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=569/H;
RX MEDLINE=87109042; PubMed=3027036;
RA Hussain M., Pastor F.I.J., Lampen J.O.;
RT "Cloning and sequencing of the blaZ gene encoding beta-lactamase III,
a lipoprotein of Bacillus cereus 569/H";
RL J. Bacteriol. 169:579-586 (1987).
CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
CC amino acid.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M15195; AAA22274.1; -
DR PIR; A27755; PNESLC.
DR HSP; P00808; 4BLM.
DR InterPro; IPR001466; Beta_lactamase.

```

DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BACTAMASE.
DR PROSITE: PS00013; PROKARYOTIC LIPOPROTEIN; 1.
DR PROSITE: PS00146; BETA_LACTAMASE_A; 1.
KW Hydrolase; Antibiotic resistance; Membrane; Lipoprotein; Signal.
FT SIGNAL 1 29
FT CHAIN 30 316 BETA-LACTAMASE III.
FT LIPID 30 30 N-ACYL DIGLYCERIDE.
FT ACT_SITE 95 95 BY SIMILARITY.
FT BINDING 257 259 SUBSTRATE (BY SIMILARITY).
SQ SEQUENCE 316 AA; 35140 MW; 27A082AB9CCBC9DF CRC64;

Query Match 1.1%; Score 8; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KPVLISIL 14
Db 278 KPVLISIL 285

RESULT 7
FDH_PSESR STANDARD; PRT; 400 AA.
AC P33160;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Formate dehydrogenase (EC 1.2.1.2) (NAD-dependent formate
DE dehydrogenase) (FDH).
OS Pseudomonas sp. (strain 101) (Achromobacter parvulus T1).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=33067;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92063805; PubMed=1954846;
RA Tishkov V.I., Galkin A.G., Egorov A.M.;
RT "NAD-dependent formate dehydrogenase of methylotrophic bacteria
RT structure.";
RL Iokl. Akad. Nauk SSSR 317:745-748(1991).
RN [2]
RP SEQUENCE OF 1-393.
RX MEDLINE=92090536; PubMed=2357236;
RA Popov V.O., Shumilin I.A., Ustinikova T.B., Lamzin V.S., Egorov T.A.;
RT "NAD-dependent formate dehydrogenase from methylotrophic bacteria
RL Bioorg. Khim. 16:324-335(1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=92283270; PubMed=1597184;
RA Lamzin V.S., Aleshin A.E., Strokopytov B.V., Yuhnevich M.G.,
RA Popov V.O., Harutyunyan E.H., Wilson K.S.;
RT "Crystal structure of NAD-dependent formate dehydrogenase.";
RL Eur. J. Biochem. 206:441-452(1992).
RN [4]
RP MUTAGENESIS OF CYS-255.
RX MEDLINE=93249485; PubMed=8484798;
RA Tishkov V.I., Galkin A.G., Marchenko G.N., Egorova O.A., Sheluho D.V.,
RA Kulakova L.B., Dementieva L.A., Egorov A.M.;
RT "Catalytic properties and stability of a Pseudomonas sp.101 formate
RT dehydrogenase mutants containing Cys-255-Ser and Cys-255-Met
RT replacements.";
RL Biochem. Biophys. Res. Commun. 192:976-981(1993).
CC -!- CATALYTIC ACTIVITY: Formate + NAD(+) = CO(2) + NADH.
CC -!- SUBUNIT: Homodimer.
CC -!- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY. STRONG, TO OTHER FDH.
DR PIR; J00334; J00334.
DR PDB; 2NAC; 26-JAN-95.
DR PDB; 2NAD; 26-JAN-95.
DR InterPro: IPR006139; 2-Hacid_DH.
DR InterPro: IPR006140; 2-Hacid_DH_C.
DR Pfam: PF00389; 2-Hacid_DH; 1.
DR Pfam: PF02826; 2-Hacid_DH_C; 1.
DR PROSITE: PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE: PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE: PS00671; D_2_HYDROXYACID_DH_3; 1.
KW Oxidoreductase; NAD; 3D-structure.
FT INIT_MET 0 0
FT DOMAIN 1 146 CATALYTIC.
FT DOMAIN 147 333 COENZYME-BINDING.
FT DOMAIN 334 400 CATALYTIC.
FT NP_BIND 192 226 NAD (ADP PART).
FT ACT_SITE 5 5 POTENTIAL.
FT ACT_SITE 284 284 SUBSTRATE BINDING.
FT MOTAGEN 255 255 C->S,M: HIGH RESISTANCE TO INACTIVATION
BY HG(2+), INCREASED STABILITY AT 25 C
AND DECREASED THERMOSTABILITY AT 45 C.
FT CONFLICT 77 77 D -> S (IN REF. 1).
FT CONFLICT 138 139 TV -> VT (IN REF. 1).
FT CONFLICT 145 145 C -> V (IN REF. 1).
FT CONFLICT 215 216 VH -> HV (IN REF. 1).
FT CONFLICT 327 327 N -> D (IN REF. 1).
FT STRAND 2 6
FT STRAND 8 8
FT TURN 12 13
FT TURN 31 32
FT TURN 46 47
FT STRAND 52 52
FT TURN 53 55
FT HELIX 56 58
FT HELIX 60 65
FT TURN 66 67
FT STRAND 69 73
FT TURN 79 80
FT HELIX 82 87
FT TURN 88 89
FT STRAND 92 96
FT TURN 97 98
FT STRAND 99 99
FT STRAND 103 103
FT HELIX 105 110
FT TURN 112 113
FT STRAND 116 119
FT TURN 125 126
FT STRAND 127 127
FT HELIX 129 134
FT TURN 135 136
FT STRAND 138 141
FT TURN 143 144
FT HELIX 147 162
FT TURN 163 164
FT HELIX 165 173
FT TURN 174 175
FT HELIX 179 183
FT TURN 184 185
FT TURN 190 191
FT STRAND 193 197
FT HELIX 201 210
FT HELIX 211 213
FT TURN 214 214
FT STRAND 216 220
FT HELIX 227 233
FT TURN 234 234
FT STRAND 235 237
FT HELIX 241 244
FT STRAND 245 247
FT STRAND 250 253
FT TURN 259 263
FT STRAND 265 265
FT TURN 267 270
FT TURN 271 272
FT TURN 275 276
FT STRAND 277 281

```

```

FT HELIX 285 287
FT STRAND 288 288
FT HELIX 289 298
FT TURN 290 298
FT TURN 299 300
FT STRAND 302 307
FT TURN 317 318
FT HELIX 320 322
FT TURN 323 323
FT TURN 325 326
FT TURN 327 327
FT STRAND 334 336
FT HELIX 338 357
FT TURN 358 358
FT HELIX 363 365
FT STRAND 366 368
FT STRAND 373 373
FT TURN 375 375
FT HELIX 376 380
SQ SEQUENCE 400 AA; 44005 MW; A354151D4312991F CRC64;

Query Match 1.1%; Score 8; DB 1; Length 400;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
Db 53 VSGELGLR 60

RESULT 8
RGA4 SCHPO STANDARD; PRT; 933 AA.
AC O74360;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable Rho-type GTPase-activating protein 4.
GN RGA4 OR SPBC28E12.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Waltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leilaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shipakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).

```

```

RN [2]
RP GENE NAME.
RX PubMed=11737264;
RA Nakano K., Mutoh T., Mabuchi I.;
RT "Characterization of GTPase-activating proteins for the function of
RT the Rho-family small GTPases in the fission yeast Schizosaccharomycetes
RT pombe.";
RL Genes Cells 6:1031-1042(2001).
CC -!- FUNCTION: GTPase-activating protein for Rho-type proteins
CC (Potential).
CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.
CC -!- SIMILARITY: Contains 1 RHO-GAP domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL031517; CAA20650.2; -.
DR GeneDB SPombe; SPBC28E12.03; -.
DR InterPro; IPR001781; LIM.
DR InterPro; IPR000198; RhoGAP.
DR Pfam; PF00412; LIM; 2.
DR Pfam; PF00620; RhoGAP; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 2.
DR SMART; SM00324; RhoGAP; 1.
DR PROSITE; PS00478; LIM DOMAIN 1; 1.
DR PROSITE; PS0023; LIM DOMAIN 2; 1.
DR PROSITE; PS0238; RHO GAP; 1.
KW GTPase activation; Repeat; LIM domain; Metal-binding; Zinc.
FT DOMAIN 22 80 LIM 1.
FT DOMAIN 81 129 LIM 2.
FT DOMAIN 753 932 RHO-GAP.
SQ SEQUENCE 933 AA; 104487 MW; 2EE75150FG2E5586 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 933;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q 206 EGLYRKSG 213
Db 779 EGLYRKSG 786
|||||||

RESULT 9
RGA2 YEAST STANDARD; PRT; 1009 AA.
AC Q06407;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Rho-type GTPase-activating protein 2.
GN RGA2 OR YDR379W OR D9481.4.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX PubMed=9169867;
RA Jacq C., Al-Moerhe J., Andre B., Arnold W., Bahr A., Ballesta J.P.G.,
RA Bagues M., Baron L., Becker A., Biteau N., Bloeker H., Bugnon C.,
RA Boskovic J., Brandt P., Brueckner M., Buitrago M.J., Coster F.,
RA Delaveau T., del Rey F., Dujon B., Eide L.G., Garcia-Cantalejo J.M.,
RA Goffeau A., Gomez-Peris A., Granotier C., Hanemann V., Hankeln T.,
RA Hoheisel J.D., Jaeger W., Jimenez A., Jonniaux J.-L., Kraemer C.,
RA Kuester H., Laamanen P., Legros Y., Louis E.J., Moeller-Ricker S.,
RA Monnet A., Moro M., Mueller-Auer S., Nussbaumer B., Paricio N.,

```

```

RA Paulin L., Perea J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,
RA Prydz H., Farnelle B., Rasmussen S.W., Renacha M., Revuelta J.L.,
RA Rieger M., Salom D., Saluz J.E., Saren A.-M., Schaefer M.,
RA Schaffe M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,
RA Urrestazu L.A., Verhaselt P., Vissers S., Voet M., Wolckaert G.,
RA Wagner G., Wambutt R., Wedler H., Woelfl S., Harris D.E.,
RA Bowman S., Brown D., Churher C.M., Connor R., Dedman K., Gentles S.,
RA Hamlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,
RA Odell C., Oliver K., Rajandream M.A., Richards C., Shore L.,
RA Walsh S.V., Barrell B.G., Dietrich F.S., Mulligan J.T., Allen E.,
RA Araujo R., Aviles E., Berno A., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Hunnicke-Smith S., Hyman R.W., Komp C.,
RA Lashkari D., Lew H., Lin D., Mosedale D., Nakahara K., Mamath A.,
RA Cefnegr P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,
RA Shogren T., Shroff N., Winant A., Telton M.A., Botstein D.,
RA Davis R.W., Johnston M., Andrews S., Brinkman R., Cooper J., Ding H.,
RA Du Z., Favell A., Fulton L., Gattung S., Greco T., Hallsworth K.,
RA Hawkins J., Hillier L., Jier M., Johnson D., Johnston L., Kirsten J.,
RA Kucaba T., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
RA Taich A., Trevisan E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R., Albermann K., Ham J., Heumann K., Kleine K.,
RA Mewes H.-W., Zollner A., Zaccaria P.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IV.";
RL Nature 387:75-78(1997).
CC -!- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHO1.
CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.
CC -!- SIMILARITY: Contains 1 RHO-GAP domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U28373; AAB64815.1; -
CC HSSP; P27986; 1PBW
CC SGD; S0002787; RGA2.
CC InterPro; IPR001781; LIM.
CC InterPro; IPR000198; RhoGAP.
CC Pfam; PF00412; LIM; 2.
CC ProDom; PD000094; RhoGAP; 1.
CC SMART; SM00132; LIM; 2.
CC SMART; SM00324; RhoGAP; 1.
CC PROSITE; PS00478; LIM DOMAIN 1; 1.
CC PROSITE; PS50023; LIM DOMAIN 2; 2.
CC PROSITE; PS50238; RHO GAP; 1.
CC GTPase activation; Repeat; LIM domain; Metal-binding; Zinc.
CC FT DOMAIN 11 68 LIM 1.
CC FT DOMAIN 69 129 LIM 2.
CC FT DOMAIN 788 1006 RHO-GAP.
CC SQ SEQUENCE 1009 AA; 113290 MW; FA13BA1DF8B35D98 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1009;
Best Local Similarity 100.0%; Pred. No.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 EGLYKSG 213
DB 822 EGLYKSG 829

RESULT 10
NU3M DROME
ID NU3M DROME STANDARD; PRT; 117 AA.
AC P18930;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).

RA Paulin L., Perea J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,
RA Prydz H., Farnelle B., Rasmussen S.W., Renacha M., Revuelta J.L.,
RA Rieger M., Salom D., Saluz J.E., Saren A.-M., Schaefer M.,
RA Schaffe M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,
RA Urrestazu L.A., Verhaselt P., Vissers S., Voet M., Wolckaert G.,
RA Wagner G., Wambutt R., Wedler H., Woelfl S., Harris D.E.,
RA Bowman S., Brown D., Churher C.M., Connor R., Dedman K., Gentles S.,
RA Hamlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,
RA Odell C., Oliver K., Rajandream M.A., Richards C., Shore L.,
RA Walsh S.V., Barrell B.G., Dietrich F.S., Mulligan J.T., Allen E.,
RA Araujo R., Aviles E., Berno A., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Hunnicke-Smith S., Hyman R.W., Komp C.,
RA Lashkari D., Lew H., Lin D., Mosedale D., Nakahara K., Mamath A.,
RA Cefnegr P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,
RA Shogren T., Shroff N., Winant A., Telton M.A., Botstein D.,
RA Davis R.W., Johnston M., Andrews S., Brinkman R., Cooper J., Ding H.,
RA Du Z., Favell A., Fulton L., Gattung S., Greco T., Hallsworth K.,
RA Hawkins J., Hillier L., Jier M., Johnson D., Johnston L., Kirsten J.,
RA Kucaba T., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
RA Taich A., Trevisan E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R., Albermann K., Ham J., Heumann K., Kleine K.,
RA Mewes H.-W., Zollner A., Zaccaria P.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IV.";
RL Nature 387:75-78(1997).
CC -!- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHO1.
CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.
CC -!- SIMILARITY: Contains 1 RHO-GAP domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U28373; AAB64815.1; -
CC HSSP; P27986; 1PBW
CC SGD; S0002787; RGA2.
CC InterPro; IPR001781; LIM.
CC InterPro; IPR000198; RhoGAP.
CC Pfam; PF00412; LIM; 2.
CC ProDom; PD000094; RhoGAP; 1.
CC SMART; SM00132; LIM; 2.
CC SMART; SM00324; RhoGAP; 1.
CC PROSITE; PS00478; LIM DOMAIN 1; 1.
CC PROSITE; PS50023; LIM DOMAIN 2; 2.
CC PROSITE; PS50238; RHO GAP; 1.
CC GTPase activation; Repeat; LIM domain; Metal-binding; Zinc.
CC FT DOMAIN 11 68 LIM 1.
CC FT DOMAIN 69 129 LIM 2.
CC FT DOMAIN 788 1006 RHO-GAP.
CC SQ SEQUENCE 1009 AA; 113290 MW; FA13BA1DF8B35D98 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 1009;
Best Local Similarity 100.0%; Pred. No.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 EGLYKSG 213
DB 822 EGLYKSG 829

RESULT 10
NU3M DROME
ID NU3M DROME STANDARD; PRT; 117 AA.
AC P18930;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).

```

```

GN MT:ND3 OR ND3.
OS Drosophila melanogaster (Fruit fly).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bretagne;
RX MEDLINE=88212147; PubMed=3130291;
RA Garesse R.;
RT "Drosophila melanogaster mitochondrial DNA: gene organization and
RL evolutionary considerations.";
RL Genetics 118:649-663(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R, and Zimbabwe 53;
RX MEDLINE=20363871; PubMed=10903372;
RA Ballard J.W.O.;
RT "Comparative genomics of mitochondrial DNA in members of the
RL Drosophila melanogaster subgroup.";
RL J. Mol. Evol. 51:48-63(2000).
RN [3]
RP SEQUENCE OF 1-29 FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=83220794; PubMed=6304652;
RA Clary D.O., Wahleithner J.A., Wolstenholme D.R.;
RT "Transfer RNA genes in Drosophila mitochondrial DNA: related 5',
RT flanking sequences and comparisons to mammalian mitochondrial tRNA
RT genes.";
RL Nucleic Acids Res. 11:2411-2425(1983).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M37275; AAA69709.1; -
CC DR EMBL; U37541; AAC47817.1; -
CC DR EMBL; AF200828; AAF77232.1; -
CC DR EMBL; AF200829; AAF77244.1; -
CC DR EMBL; J01404; AAB59244.1; -
CC DR PIR; S01185; S01185.
CC DR FlyBase; FBgn0013681; mt:ND3.
CC DR InterPro; IPR000440; Oxidored_q4.
CC DR Pfam; PF00507; oxidored_q4; 1.
CC DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
CC KW SEQUENCE 117 AA; 13563 MW; 6DB7AFDA72924892 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No.19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 KALIDRE 467
DB 28 KALIDRE 34

RESULT 11
NU3M DROSU
ID NU3M DROSU STANDARD; PRT; 117 AA.
AC P51940;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).
GN MT:ND3 OR ND3.

```

```

OS Drosophila subobscura (Fruit fly).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7241;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSP. TUE 3; PubMed=1454843;
RX MEDLINE=3087561; PubMed=1454843;
RA Volz-Lingenhoel A., Solignac M., Sperlich D.;
RT "Stable heteroplasm for a large-scale deletion in the coding region
of Drosophila subobscura mitochondrial DNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11528-11532(1992).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65129; CAA46258.1; -.
DR FLYBase; FBgn0012957; Dsub\mt.ND3.
DR InterPro; IPR000440; Oxidored_q4.
DR Pfam; PF00507; oxidored_q4; 1.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 117 AA; 13427 MW; 01FEDA1E94FB4012 CRC64;
-----
Query Match 0.9%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred.No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 461 KALIDRE 467
Db 28 KALIDRE 34
|||||

RESULT 12
ID NU3M DROYA STANDARD; PRT; 117 AA.
AC P07705;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3).
GN ND3.
OS Drosophila yakuba (Fruit fly).
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2317.6 Ivory Coast;
RX MEDLINE=86089137; PubMed=3001325;
RA Clary D.O., Wolstenholme D.R.;
RT "The mitochondrial DNA molecular of Drosophila yakuba: nucleotide
sequence, gene organization, and genetic code.";
RL J. Mol. Evol. 22:252-271(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2317.6 Ivory Coast;
RX MEDLINE=84169539; PubMed=6324117;
RA Clary D.O., Wolstenholme D.R.;
RT "A cluster of six tRNA genes in Drosophila mitochondrial DNA that
includes a gene for an unusual tRNASerAGY.";
RL Nucleic Acids Res. 12:2367-2379(1984).
RN [4]
RP -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

```

```

CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03240; CAA26991.1; -.
DR FLYBase; FBgn0013204; Dyak\mt.ND3.
DR InterPro; IPR000440; Oxidored_q4.
DR Pfam; PF00507; oxidored_q4; 1.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 117 AA; 13457 MW; 3291F89BE5DD1D19D CRC64;
-----
Query Match 0.9%; Score 7; DB 1; Length 117;
Best Local Similarity 100.0%; Pred.No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 461 KALIDRE 467
Db 28 KALIDRE 34
|||||

RESULT 13
ID PYRI_ECOLI STANDARD; PRT; 152 AA.
AC P00478;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aspartate carbamoyltransferase regulatory chain.
GN PYRI OR B4244 OR C5344 OR Z5855 OR ECS5221.
OS Escherichia coli,
OS Escherichia coli O6, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84119419; PubMed=6364131;
RA Schachman H.K., Pauza C.D., Navre M., Karels M.J., Wl L., Yang Y.R.;
RT "Location of amino acid alterations in mutants of aspartate
transcarbamoylase: structural aspects of interallelic
complementation.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:115-119(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Maynew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [4]
RP SEQUENCE FROM N.A.

```

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
RL Nature 409:529-533 (2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:111-22 (2001).
RN [16]
RP SEQUENCE OF 127-152 FROM N.A.
RX MEDLINE=86143826; PubMed=3912513;
RA Cunin R., Jacobs A., Charlier D.R.M., Crabeel M., Herve G.,
RA Glandsdorff N., Pierard A.;
RT "Structure-function relationship in allosteric aspartate
RT carbamoyltransferase from *Escherichia coli*. I. Primary structure of a
RT pyri gene encoding a modified regulatory subunit.";
RL J. Mol. Biol. 186:707-713 (1985).
RN [7]
RP PRELIMINARY SEQUENCE.
RX MEDLINE=68284659; PubMed=4872216;
RA Weber K.;
RT "New structural model of *E. coli* aspartate transcarbamylase and the
RT amino-acid sequence of the regulatory polypeptide chain.";
RL Nature 218:1116-1119 (1968).
RN [8]
RP SEQUENCE OF 1-10.
RC STRAIN=K12 / W3110;
RA Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
RA Frutiger S., Faquet N., Wilkins M., Appel R.D., Bairoch A.,
RA Hochstrasser D.F.;
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
RN [9]
RP SEQUENCE OF 1-12.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of *Escherichia coli* K-12.";
RL Electrophoresis 18:1259-1313 (1997).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=79074799; PubMed=364472;
RA Monaco H.L., Crawford J.L., Lipscomb W.N.;
RT "Three-dimensional structures of aspartate carbamoyltransferase from
RT *Escherichia coli* and of its complex with cytidine triphosphate.";
RL Proc. Natl. Acad. Sci. U.S.A. 75:5276-5280 (1978).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND SUBUNITS.
RX MEDLINE=84248054; PubMed=6377306;
RA Ke H., Honzatko R.B., Lipscomb W.N.;
RT "Structure of unligated aspartate carbamoyltransferase of *Escherichia*
RT coli at 2.6-A resolution.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:4037-4040 (1984).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=91104727; PubMed=2271528;
RA Stevens R.C., Gouaux J.E., Lipscomb W.N.;
RT "Structural consequences of effector binding to the T state of
RT aspartate carbamoyltransferase: crystal structures of the unligated
RT and ATP- and CTP-complexed enzymes at 2.6-A resolution.";
RL Biochemistry 29:7702-7715 (1990).
RN [13]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=91104728; PubMed=2271529;
RA Gouaux J.E., Stevens R.C., Lipscomb W.N.;
RT "Crystal structures of aspartate carbamoyltransferase ligated with
RT phosphonoacetamide, malonate, and CTP or ATP at 2.8-A resolution and
RT neutral pH.";
RL Biochemistry 29:7702-7715 (1990).
RN [14]
RP FUNCTION: INVOLVED IN ALLOSTERIC REGULATION OF ASPARTATE
RX CARBAMOYLTRANSFERASE.
CC COFACTOR: Binds 1 zinc ion per subunit.
CC SUBUNIT: HETERODODECAMER (2C3:3R2) OF SIX CATALYTIC PYRB CHAINS
CC ORGANIZED AS TWO TRIMERS (C3), AND SIX REGULATORY PYRI CHAINS
CC ORGANIZED AS THREE DIMERS (R2).
CC SIMILARITY: BELONGS TO THE PYRI FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
RN [15]
RP EMBL; K01472; AAA24477.1; -;
DR EMBL; U14003; AAA97141.1; -;
DR EMBL; AF000495; AAC77201.1; -;
DR EMBL; AB016771; AAN83765.1; -;
DR EMBL; AF005657; BAG59442.1; -;
DR EMBL; AF002568; BAB38644.1; -;
DR EMBL; M28578; AAA24487.1; -;
DR PIR; A93985; DTECR.
DR PIR; E91281; E91281.
DR PIR; F86122; F86122.
DR PDB; 1ACM; 15-JUL-92.
DR PDB; 2ATC; 15-JUL-92.
DR PDB; 8ATC; 15-OCT-90.
DR PDB; 1AT1; 15-OCT-90.
DR PDB; 2AT1; 15-OCT-90.
DR PDB; 3AT1; 15-OCT-90.
DR PDB; 4AT1; 15-OCT-90.
DR PDB; 5AT1; 31-JUL-94.
DR PDB; 6AT1; 15-OCT-90.
DR PDB; 7AT1; 15-OCT-90.
DR PDB; 8AT1; 31-JUL-94.
DR PDB; 1RAA; 31-JAN-94.
DR PDB; 1RAB; 31-JAN-94.
DR PDB; 1RAC; 31-JAN-94.
DR PDB; 1RAD; 31-JAN-94.
DR PDB; 1RAE; 31-JAN-94.
DR PDB; 1RAF; 31-JAN-94.
DR PDB; 1RAG; 31-JAN-94.
DR PDB; 1RAH; 31-JAN-94.
DR PDB; 1RAI; 31-JAN-94.
DR PDB; 1RBE; 14-OCT-98.
DR PDB; 1D09; 23-MAY-00.
DR PDB; 1EZZ; 22-NOV-00.
DR PDB; 1F1B; 22-NOV-00.
DR PDB; 1I50; 02-MAY-01.
DR SWISS-2DPAGE; P00478; COLI.
DR ECODDBASE; I017.5; 6TH EDITION.
DR Ecogene; EGI0811; pyri.
DR HAMAP; MF_00002; -; 1.
DR InterPro; IPR002801; Pyri.
DR Pfam; PF01948; Pyri; 1.
DR Pfam; PF02748; Pyri_C; 1.
DR ProDom; PD006194; Pyri; 1.
DR TIGRFAMs; TIGR00340; AtCase reg; 1.
KW Pyrimidine biosynthesis; Zinc; 3D-structure; Complete proteome.
FT INIT MET 0
FT METAL 108 108 ZINC.
FT METAL 113 113 ZINC.


```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05641; CAA29130.1; -.
DR EMBL; AE008909; AAL23278.1; -.
DR PIR; S00050; DTBECT.
DR HSP; P00478; IRAB.
DR StyGene; SG10329; PYRI.
DR HAMAP; MF_00002; -.
DR InterPro; IPR002801; PYRI.
DR Pfam; PF01948; PYRI; 1.
DR Pfam; PF02748; PYRI; C; 1.
DR ProDom; PD006194; PYRI; 1.
DR TIGRFAMs; TIGR00240; AtCase reg; 1.
KW Pyrimidine biosynthesis; Zinc; Complete proteome.
FT INIT MET 0
FT METAL 108 108 ZINC (BY SIMILARITY).
FT METAL 113 113 ZINC (BY SIMILARITY).
FT METAL 137 137 ZINC (BY SIMILARITY).
FT METAL 140 140 ZINC (BY SIMILARITY).
SQ SEQUENCE 152 AA; 16955 MW; BCEAED84E045B4BB CRC64;

Query Match 0.9%; Score 7; DB 1; Length 152;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 VVGKSRP 44
Db 90 VVGKSRP 96
|||||

RESULT 16
Y004_BPL2
ID Y004_BPL2 STANDARD; PRT; 161 AA.
AC P42539;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Hypothetical 18.2 kDa protein (ORF4).
OS Bacteriophage L2.
OC Viruses; dsDNA viruses, no RNA stage; Plasmaviridae; Plasmavirus.
OX NCBI_TaxID=46014;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94215869; PubMed=8163159;
RA Manioff J., Kampo G.J., Dascher C.C.;
RT "Sequence analysis of a unique temperature phase: mycoplasma virus L2."
RL Gene 141:1-8(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13696; AAA87960.1; -.
DR Hypothetical protein.
SQ SEQUENCE 161 AA; 18225 MW; 7FCE715B0AAE2715 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 161;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 DDAHAHA 321
Db 151 DDAHAHA 157
|||||

RESULT 17
ATP4_IPOBA
ID ATP4_IPOBA STANDARD; PRT; 200 AA.

```

```

AC Q40089;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase delta' chain, mitochondrial precursor (EC 3.6.3.14).
OS Ipomoea batatas (Sweet potato) (Batate).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=4120;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9211272; PubMed=1370454;
RA Morikawa A., Aiso K., Asahi T., Nakamura K.;
RT "The delta'-subunit of higher plant six-subunit mitochondrial
RT F1-ATPase is homologous to the delta-subunit of animal mitochondrial
RL J. Biol. Chem. 267:72-76(1992).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- SIMILARITY: Belongs to the ATPase epsilon chain family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10660; BAA01511.1; -.
DR PIR; A41740; A41740.
DR InterPro; IPR001469; ATPsynth DE.
DR Pfam; PF02823; ATP-synt DE_N; 1.
DR ProDom; PD000944; ATPsynth_DE; 1.
DR TIGRFAMs; TIGR01216; ATP_synth_epsilon; 1.
KV ATP synthesis; CF(1); Hydrogen ion transport; Hydrolase;
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 21 MITOCHONDRION
FT CHAIN 22 200 ATP SYNTHASE DELTA' CHAIN.
SQ SEQUENCE 200 AA; 21318 MW; 1563325C81BCE828 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 DRIDANL 680
Db 154 DRIDANL 160
|||||

RESULT 18
DEOC_HALN1
ID DEOC_HALN1 STANDARD; PRT; 212 AA.
AC Q9HP08;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Probable deoxyribose-phosphate aldolase (EC 4.1.2.4)
DE (Phosphodeoxyriboaldolase) (Deoxyriboaldolase) (DERA).
GN DEOC OR VNGI859G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W., Dale H.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Jung K.-H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-
CC glycerinaldehyde 3-phosphate + acetaldehyde.
CC -!- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES. DEOC
CC SUBFAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE005086; AAG20062.1; -.
DR PIR; B84337; B84337.
DR HAMAP; MF 00114; -.
DR InterPro; IPR002915; Deoc.
DR InterPro; IPR003009; FMN_enzyme.
DR Pfam; PF01791; Deoc; 1.
KW Lyase; Schiff base; Complete proteome.
FT BINDING 151 151 SCHIFF-BASE WITH ACETALDEHYDE (BY
FT SIMILARITY).
FT ACT SITE 176 176 BY SIMILARITY.
SQ SEQUENCE 212 AA; 21472 MW; F9DCF56817DE87DA CRC64;

Query Match 0.9%; Score 7; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 389 FIGRLKG 395
DB 94 FIGRLKG 100
|||||
|||||

RESULT 19
Y116 METJA STANDARD; PRT; 225 AA.
AC Q57580;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0116.
GN MJ0116.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

```

```

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073 (1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67469; AAB98097.1; -.
DR PIR; D64314; D64314.
DR TIGR; MJ0116; -.
DR InterPro; IPR003785; Creatininase.
DR Pfam; PF02633; Creatininase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 225 AA; 25305 MW; 929EF2CEC84DE35C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 LRLSSGN 184
DB 3 LRLSSGN 9
|||||
|||||

RESULT 20
FA3A HUMAN STANDARD; PRT; 230 AA.
ID FA3A HUMAN
AC P98173; Q96H51;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein FAM3A precursor (2-19 protein).
GN FAM3A OR 2-19 OR 2.19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=94068527; PubMed=8248200;
RA Bione S., Tamani F., Maestrini E., Tribioli C., Poustka A.,
RA Torri G., Rivella S., Toniolo D.;
RT "Transcriptional organization of a 450-kb region of the human X
RT chromosome in Xq28.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10977-10981 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Zollo M., Mazzarella R., Bione S., Toniolo D., Schlessinger D.,
RA D'Urso M., Chen E.Y.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96311563; PubMed=8733135;
RA Chen E.Y., Zollo M., Mazzarella R.A., Ciccodicola A., Chen C.-N.,
RA Zuo L., Heiner C., Burrough F.W., Ripetto M., Schlessinger D.,
RA D'Urso M.;
RT "Long-range sequence analysis in Xq28: thirteen known and six
RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
RT G6PD loci.";
RL Hum. Mol. Genet. 5:659-668 (1996).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANT TRP-213.
RX TISSUE=Pancreas;
RX MEDLINE=2388257; PubMed=12477932;
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

```

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=22150867; PubMed=12160727;
RA Zhu Y., Xu G., Patel A., McLaughlin M.M., Silverman C., Knecht K.A.,
RA Sweitzer S., Li X., McDonnell P., Mirabile R., Zimmerman D., Boyce R.,
RA Tierney L.A., Hu E., Livi G.P., Wolf B.A., Abdel-Meguid S.S.,
RA Rose G.D., Aurora R., Hensley P., Briggs M., Young P.R.,
RT "Cloning, expression, and initial characterization of a novel
RT cytokine-like gene family."
RL Genomics 80:144-150(2002).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- TISSUE SPECIFICITY: IN SIMILAR AMOUNTS IN TESTIS, PANCREAS,
CC ADRENAL, PLACENTA, BRAIN, FETAL BRAIN, LIVER, KIDNEY, SKELETAL
CC MUSCLE, HEART.
CC -!- SIMILARITY: BELONGS TO THE FAM3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL; X55448; CAA39090.1; -;
CC DR EMBL; X87193; CAA60645.1; -;
CC DR EMBL; L44140; AAA92652.1; -;
CC DR EMBL; BC008912; AB08912.1; -;
CC DR EIR; 137095; 137095.
CC DR Genew; HGNC:13749; FAM3A.
CC KW Signal; Polymorphism.
CC FT SIGNAL 1 33 POTENTIAL.
CC FT CHAIN 34 230 PROTEIN FAM3A.
CC FT DISULFID 59 222 POTENTIAL.
CC FT DISULFID 65 87 POTENTIAL.
CC FT VARIANT 213 213 C->W (IN dbSNP:1050788).
CC FT /FTID=VAR_011923.
CC SQ SEQUENCE 230 AA; 25069 MW; FE3934D91F98CAAD CRC64;

Query Match 0.9%; Score 7; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 NGVSGEL 176
Db 111 NGVSGEL 117

RESULT 21
FAJA_MOUSE
ID FA3A_MOUSE STANDARD; PRT; 230 AA.
AC Q9D8T0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein FAM3A precursor.

GN FAM3A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- SIMILARITY: BELONGS TO THE FAM3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL; AK007715; BAB25208.1; -;
CC DR MGD; MGI:1913544; 1810037C20Rik.
CC KW Signal.
CC FT SIGNAL 1 33 POTENTIAL.
CC FT CHAIN 34 230 PROTEIN FAM3A.
CC FT DISULFID 59 222 POTENTIAL.
CC FT DISULFID 65 87 POTENTIAL.
CC SQ SEQUENCE 230 AA; 25393 MW; 6239418ED8258087 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 NGVSGEL 176
Db 111 NGVSGEL 117

RESULT 22
KCX_STRCO STANDARD; PRT; 231 AA.
AC Q9EW6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytidylate kinase (EC 2.7.4.14) (CK) (Cytidine monophosphate kinase)
DE (CMP kinase).
GN CMK OR SCOL760 OR 2SCI34.13C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]

```

SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145; PubMed=12000953;
RX MEDLINE=21996410;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -I- CATALYTIC ACTIVITY: ATP + (d)CMP = ADP + (d)CDP.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE CYTIDYLATE KINASE FAMILY. SUBFAMILY 1.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL939110; CAC12933.1; -
DR HSSP; P23863; 1CKE.
DR HAMAP; MF_00238; 1.
DR InterPro; IPR003136; Cytidylate_kin.
DR Pfam; PF02224; Cytidylate_kin; 1.
DR TIGRFAMs; TIGR00017; cmk; 1.
DR Transferase; Kinase; ATP-binding; Complete proteome.
FT NP BIND 18 26 ATP (BY SIMILARITY).
FT SEQUENCE 231 AA; 23600 MW; 76AC463B6F5BF7D7E CRC64;

Query Match 0.9%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 KAGDAVE 371
Db 199 KAGDAVE 205

RESULT 23
ID ULA3 HCMVA STANDARD; PRT; 249 AA.
AC P16734; 1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE Protein UL103.
GN UL103.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,
RA Reddell E., Satchwell S.C., Tomlinson P., Weston K.M., Barrall B.G.;
RA "Analysis of the protein-coding content of the sequence of human
RA cytomegalovirus strain AD169.";
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -I- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL7,
CC EHV-1 55, VZV 53, EBV BBRF2, HCMV UL103 AND HVS-1 42.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17403; CAA35339.1; -
DR PIR; S09868; S09868.
DR InterPro; IPR002600; Herpes UL7.
DR Pfam; PF01677; Herpes UL7; 1.
DR SEQUENCE 249 AA; 28636 MW; A8D9F8F9F02FE9D CRC64;

Query Match 0.9%; Score 7; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 723 VKADNLL 729
Db 95 VKADNLL 101

RESULT 24
ID YE94_XYLFA STANDARD; PRT; 281 AA.
AC Q9PD85;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein XF1494.
GN XF1494.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferio J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Vertovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000)
CC -I- SIMILARITY: BELONGS TO THE UPF0162 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB003979; AAF84303.1; ALT_INIT.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 281 AA; 32189 MW; 47D8879D22181684 CRC64;

Query Match          0.9%; Score 7; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 RYLQLYP 292
DB 252 RYLQLYP 258

RESULT 25
HEM3_RHILO
ID HEM3_RHILO STANDARD; PRT; 308 AA.
AC Q98E17;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Porphobilinogen deaminase (EC 4.3.1.8) (PBG) (Hydroxymethylbilane
DE synthase) (HMB) (Pre-uroporphyrinogen synthase).
GN HEMC OR MLL4223.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
CC -I- FUNCTION: Tetrapolymerization of the monopyrrole PBG into the
CC hydroxymethylbilane preuroporphyrinogen in several discrete steps.
CC -I- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O =
CC hydroxymethylbilane + 4 NH(3).
CC -I- COFACTOR: Covalently binds a dipyrromethane cofactor to which the
CC porphobilinogen subunits are added (By similarity).
CC -I- PATHWAY: Porphyrin biosynthesis by the C5 pathway; fourth step.
CC -I- SUBUNIT: Monomer (By similarity).
CC -I- SIMILARITY: BELONGS TO THE HMB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AP003003; BAB50931.1; -.
DR HAMAP; MF_00260; 1.
DR InterPro; IPR00860; Porphobil_deam.
DR Pfam; PF01379; Porphobil_deam; 1.
DR Pfam; PF03900; Porphobil_deamC; 1.
DR PRINTS; PR00151; PORPHEMNAE.
DR ProDom; PD002745; Porphobil_deam; 1.
DR TIGRfams; TIGR00212; hemC; 1.
DR PROSITE; PS00533; PORPHOBILINOGEN DEAM; 1.
KW Porphyrin biosynthesis; Lyase; Complete proteome.
FT BINDING 243 243 PYRROMETHANE COFACTOR (BY SIMILARITY).
```

```
SQ SEQUENCE 308 AA; 32579 MW; 8164DE3A9FBE442 CRC64;

Query Match          0.9%; Score 7; DB 1; Length 308;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LAQAHEH 27
DB 16 LAQAHEH 22

RESULT 26
TPIC_ARATH
ID TPIC_ARATH STANDARD; PRT; 315 AA.
AC Q9SKF6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Triosephosphate isomerase, chloroplast precursor (EC 5.3.1.1) (TIM).
GN TIM OR AT2G21170 OR F26H11.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shih M.-C.;
RT "Structure and regulation of nuclear genes encoding chloroplast and
RT cytosolic triosephosphate isomerase in Arabidopsis thaliana.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=2083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Bisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
CC -I- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone
CC phosphate.
CC -I- PATHWAY: Calvin cycle.
CC -I- SUBUNIT: Homodimer (By similarity).
CC -I- SUBCELLULAR LOCATION: Chloroplast (By similarity).
CC -I- MISCELLANEOUS: IN PLANTS, THERE ARE TWO TYPES OF TPIS, CYTOSOLIC
CC AND PLASTID.
CC -I- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AC006264; AAD29799.1; -.
DR EMBL; AF247559; AAF70259.1; -.
DR PIR; A84598; A84598.
DR HSP; P00940; ITPH.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
DR TIGRfams; TIGR00419; tim; 1.
DR PROSITE; PS00171; TIM; 1.
KW Isomerase; Chloroplast; Transit peptide; Calvin cycle.
```

```
FT TRANSIT 1 60 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 61 315 TRIOSPHOSPHATE ISOMERASE.
FT ACT SITE 155 155 BY SIMILARITY.
FT ACT SITE 225 225 BY SIMILARITY.
SQ SEQUENCE 315 AA; 33345 MW; BD580F9B37FA0269 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 671 SLTDRID 677
    |||||
Db 114 SLTDRID 120

RESULT 27
FLGJ VIBCH STANDARD; PRT; 334 AA.
AC Q9KQ15; O30855;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peptidoglycan hydrolase flgJ (EC 3.2.1.-) (Muramidase flgJ).
GN FLGJ OR VC2192.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor Inaba V86;
RX MEDLINE=98415119; PubMed=9742694;
RA Das M., Chopra A.K., Wood T., Peterson J.W.;
RT "Cloning, sequencing and expression of the flagellin core protein and
RT other genes encoding structural proteins of the Vibrio cholerae
RT flagellum".
RL FEMS Microbiol. Lett. 165:239-246 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.B., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae".
RL Nature 406:477-483 (2000).
CC -!- FUNCTION: FLAGELLUM-SPECIFIC MURAMIDASE WHICH HYDROLYZES THE
CC PEPTIDOGLYCAN LAYER TO ASSEMBLE THE ROD STRUCTURE IN THE
CC PERIPLASMIC SPACE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- MISCELLANEOUS: PROBABLY EXPORTED VIA THE FLAGELLUM-SPECIFIC EXPORT
CC PATHWAY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC GLYCOSYL HYDROLASES.
CC -!- SIMILARITY: BELONGS TO THE FLGJ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF019213; AAC33157.1; --
CC EMBL; AE004291; AAF95337.1; --
CC FIC; G82105; G82105.
CC TIGR; VC2192; --
```

```
DR InterPro; IPR002901; Amidase 4.
DR InterPro; IPR00423; Flag FlgJ.
DR Pfam; PF01832; Amidase 4; 1.
DR PRINTS; PR01002; FLGFLGJ.
DR SMART; SM00047; LY22; 1.
KW Flagella; Hydrolase; Glycosidase; Cell wall; Periplasmic;
KW Complete proteome.
FT DOMAIN 160 334 CATALYTIC.
FT ACT SITE 231 231 POTENTIAL.
FT ACT SITE 256 256 POTENTIAL.
FT CONFLICT 90 90 S -> C (IN REF. 1).
FT CONFLICT 312 334 YKSPAGIFRRSFSSIFSAYQS -> L (IN REF. 2).
SQ SEQUENCE 334 AA; 37271 MW; DEA90DB0745450A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LRQKAVN 72
    |||||
Db 20 LRQKAVN 26

RESULT 28
XYNB STRLI STANDARD; PRT; 335 AA.
ID XYNB STRLI STANDARD; PRT; 335 AA.
AC P26515;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Endo-1,4-beta-xylanase B precursor (EC 3.2.1.8) (Xylanase B)
DE (1,4-beta-D-xylan xylanohydrolase B).
GN XLNB.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 42-72.
RC STRAIN=66 / 1326;
RX MEDLINE=92077439; PubMed=1743521;
RA Shareck F., Roy C., Yaguchi M., Morosoli R., Kluepfel D.;
RT "Sequences of three genes specifying xylanases in Streptomyces
RT lividans".
RL Gene 107:75-82 (1991).
RN [2]
RP REVISIONS TO 29-32 AND 252-307.
RC STRAIN=66 / 1326; PubMed=7533741;
RX MEDLINE=95189090; PubMed=7533741;
RA Shareck F., Biely P., Morosoli R., Kluepfel D.;
RT "Analysis of DNA flanking the xlnB locus of Streptomyces lividans
RT reveals genes encoding acetyl xylan esterase and the RNA component of
RT ribonuclease P".
RN Gene 153:105-109 (1995).
RN [3]
RP REVISION TO 225.
RA Shareck F.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Contributes to hydrolyze hemicellulose, the major
CC component of plant cell-walls. XLNA and XLNB seem to act
CC sequentially on the substrate to yield xylobiose and xylose
CC as carbon sources.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC linkages in xylans.
CC -!- PATHWAY: Xylan degradation.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY G (FAMILY 11 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M64552; AAC06114.2; -.
DR HSP; P09850; IKNB.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR001137; Glyco_hydro_11.
DR Pfam; PF00457; Glyco_hydro_11; 1.
DR PRINTS; PR00911; GLHYDRLASE11.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
KW Xylan degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 41
FT CHAIN 42 335
FT DOMAIN 42 230 ENDO-1,4-BETA-XYLANASE B.
FT DOMAIN 231 249 CATALYTIC.
FT DOMAIN 250 335 LINKER ("HINGE") (GLY-RICH BOX).
FT ACT_SITE 128 128 XYLAN-BINDING (POTENTIAL).
FT ACT_SITE 128 128 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 218 218 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 335 AA; 35575 MW; 513B1458BF0FF0CF CRC64;

Query Match 0.9%; Score 7; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 RGPVTL 146
Db 12 RGPVTL 18

RESULT 29
TRML_ARCFU STANDARD; PRT; 349 AA.
ID TRML_ARCFU
AC Q29443;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N(2),N(2)-dimethylguanosine tRNA methyltransferase (EC 2.1.1.32)
DE (tRNA(Guanine-26,N(2)-N(2)) methyltransferase) (tRNA 2,2-
dimethylguanosine-26 methyltransferase)
DE (tRNA(m(2,2)G26)dimethyltransferase).
GN TRM1 OR AF0815.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-P., White O., Nelson K.E.,
Retchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
Richardson D.L., Kervagge A.R., Graham D.E., Kyripides N.C.,
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glöckle A., Zhou L.,
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Dimethylates a single guanine residue at position 26 of
a number of tRNAs using S-adenosyl-L-methionine as donor of the
methyl groups (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
homocysteine + tRNA containing N(2)-methylguanine.
CC -1- SIMILARITY: BELONGS TO THE TRM1 FAMILY.
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE001048; AAB90427.1; -.
DR PIR; G69351; G69351.
DR TIGR; AF0815; -.
DR HAMAP; MF 00290; -. 1.
DR InterPro; IPR002905; TRM.
DR Pfam; PF02005; TRM; 1.
DR TIGRFAMs; TIGR00308; TRM1; 1.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
SQ SEQUENCE 349 AA; 38808 MW; 963859AFC18987F9 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 ENGVSGE 175
Db 81 ENGVSGE 87

RESULT 30
MLTA_VIBCH STANDARD; PRT; 368 AA.
ID MLTA_VIBCH
AC Q9KQ4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Membrane-bound lytic murein transglycosylase A precursor (EC 3.2.1.-)
DE (Murein hydrolase A).
GN MLTA OR VC2312.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae."
RL Nature 406:477-483(2000).
CC -1- FUNCTION: MUREIN-DEGRADING ENZYME. MAY PLAY A ROLE IN RECYCLING OF
MUREIN PEPTIDES DURING CELL ELONGATION AND/OR CELL DIVISION. DEGRADES
MUREIN GLYCAN STRANDS AND INSOLUBLE, HIGH-MOLECULAR WEIGHT MUREIN
SACCHARI (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Cleavage of the beta-1,4-glycosidic bond
between N-acetylmuramic acid and N-acetylglucosamine residues,
thereby conserving the energy in a newly synthesized
1,6-anhydrobond in the muramic acid residue.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
anchor (By similarity).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```



```
CC ENBL; AE004301; AAF95456.1; -.
DR DR PIR; A82093; A82093.
DR TIGR; VC2312; -.
DR InterPro; IPR005300; Mita.
DR Pfam; PF03562; Mita; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Cell wall; Hydrolase; Glycosidase; Signal; Lipoprotein;
FT Outer membrane; Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 368
MEMBRANE-BOUND LYTIC MUREIN
TRANSGLYCOSYLASE A. (PROBABLE).
FT LIPID 17 17 N-ACYL DIGLYCERIDE (PROBABLE).
FT SEQUENCE 368 AA; 40458 MW; C1341A9D21A41DBB CRC64;

Query Match 0.9%; Score 7; DB 1; Length 368;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 383 IELRHQP 389
Db 121 IELRHQP 127
|||||

RESULT 31
HEM3_PEA STANDARD; PRT; 369 AA.
ID HEM3_PEA
AC Q43082;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Porphobilinogen deaminase, chloroplast precursor (EC 4.3.1.8) (PBG)
DE (Hydroxymethylbilane synthase) (HMS) (Pre-uroporphyrinogen synthase).
GN HEMC.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=cv. Little Marvel; TISSUE=Leaf;
RX MEDLINE=94269188; PubMed=7516080;
RA Witty M., Wallace-Cook A.D.M., Albrecht H., Spano A.J., Michel H.,
RA Shabanowitz J., Hunt D.F., Tinko M.P., Smith A.G.;
RT "Structure and expression of chloroplast-localized porphobilinogen
RT deaminase from pea (Pisum sativum L.) isolated by redundant
RT polymerase chain reaction."
RL Plant Physiol. 103:139-147(1993).
CC -!- FUNCTION: TETRAPOLYMERIZATION OF THE MONOPYRROLE PBG INTO THE
CC -!- HYDROXYMETHYLBILANE PREUROPORPHYRINOGEN IN SEVERAL DISCRETE STEPS.
CC -!- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O =
CC hydroxymethylbilane + 4 NH(3).
CC -!- COFACTOR: COVALENTLY BINDS A DIPYRROMETHANE COFACTOR TO WHICH
CC THE PORPHOBILINOGEN SUBUNITS ARE ADDED.
CC -!- PATHWAY: porphyrin biosynthesis by the C5 pathway; fourth step.
CC Involved in chlorophyll biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE HMS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; X73418; CAA51820.1; -.
DR PIR; S35873; JQ2278.
DR HSPF; P06983; 2YFN.
DR InterPro; IPR000860; Porphobil deam.
DR Pfam; PF01379; Porphobil deam; 1.
DR Pfam; PF03900; Porphobil deamC; 1.
```

```
DR PRINTS; PR00151; PORPHBDMWASE.
DR ProDom; PD002745; Porphobil deam; 1.
DR TIGRFAMS; TIGR00212; hemC; 1.
DR PROSITE; PS00533; PORPHOBILINOGEN DEAM; 1.
KW Porphyrin biosynthesis; Chlorophyll biosynthesis; Lyase; Chloroplast;
KW Transit peptide.
FT TRANSIT 1 46 CHLOROPLAST.
FT CHAIN 47 369 PORPHOBILINOGEN DEAMINASE.
FT BINDING 303 303 PYRROMETHANE COFACTOR (BY SIMILARITY).
FT SEQUENCE 369 AA; 39930 MW; F7660D8390894431 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 369;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 21 LAQAHEH 27
Db 73 LAQAHEH 79
|||||

RESULT 32
ALR_MYCLE STANDARD; PRT; 388 AA.
ID ALR_MYCLE
AC P38056;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanine racemase (EC 5.1.1.1).
GN ALR OR ML0375 OR B229 C3_243.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: Provides the D-alanine required for cell wall
CC biosynthesis (By similarity).
CC -!- CATALYTIC ACTIVITY: L-alanine = D-alanine.
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- PATHWAY: D-alanine branch of peptidoglycan biosynthesis; first
CC step.
CC
CC -!- SIMILARITY: Belongs to the alanine racemase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U00020; AAA17309.1; -.
DR EMBL; AL583918; CAC29883.1; -.
DR PIR; S72995; S72995.
DR HSPF; P10724; 1BD0.
DR Leproma; ML0375; -.
```

```

DR HAMAP; MF_01201; -. 1.
DR InterPro; IPR000821; Ala_racemase.
DR Pfam; PF00842; Ala_racemase_C; 1.
DR Pfam; PF01168; Ala_racemase_N; 1.
DR PRINTS; PR00992; ALARACEMASE.
DR TIGRFAMs; TIGR00492; alr; 1.
DR PROSITE; PS00395; ALANINE_RACEMASE; 1.
KW Isomerase; Pyridoxal phosphate; Cell wall; Peptidoglycan synthesis;
KW Complete proteome.
FT ACT_SITE 44 44 CATALYTIC BASE SPECIFIC TO D-ALANINE (BY
FT ACT_SITE 44 44 SIMILARITY).
FT ACT_SITE 273 273 CATALYTIC BASE SPECIFIC TO L-ALANINE (BY
FT ACT_SITE 273 273 SIMILARITY).
FT BINDING 44 44 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 388 AA; 41083 MW; 84075B683D00D196 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LLAQAE 26
Db 196 LLAQAE 202

RESULT 33
PEPF RABIT
ID PEPF RABIT STANDARD; PRT; 388 AA.
AC P27823;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pepsin F precursor (EC 3.4.23.1).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koizumi O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -1- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: hydrophobic, preferably
CC aromatic, residues in P1 and P1' positions. Cleaves 1-Phe-|-Val-2,
CC 4-Gln-|-His-5, 13-Glu-|-Ala-14, 14-Ala-|-Leu-15, 15-Leu-|-Tyr-16,
CC 16-Tyr-|-Leu-17, 23-Gly-|-Phe-24, 24-Phe-|-Phe-25 and 25-Phe-|-
CC Tyr-26 bonds in the B chain of insulin.
CC -1- DEVELOPMENTAL STAGE: EARLY POSTNATAL.
CC -1- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M59238; AAA31440.1; -.
CC HSP; P00794; 4CMS.
CC MEROPS; A01.051; -.
DR InterPro; IPR001969; Asprotease_site.
DR InterPro; IPR001461; AsproteaseA1.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.

```

```

DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 15 ACTIVATION PEPTIDE.
FT PROPEP 16 58 PEPSIN F.
FT CHAIN 59 388 BY SIMILARITY.
FT ACT_SITE 92 92 BY SIMILARITY.
FT ACT_SITE 275 275 BY SIMILARITY.
FT DISULFID 105 110 BY SIMILARITY.
FT DISULFID 266 270 BY SIMILARITY.
FT DISULFID 309 343 BY SIMILARITY.
SQ SEQUENCE 388 AA; 42786 MW; 24792BE393594B3A CRC64;

Query Match 0.9%; Score 7; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLVDVA 159
Db 152 AGLVDVA 158

RESULT 34
A23D DROME
ID A23D DROME STANDARD; PRT; 398 AA.
AC Q24093; Q9VOM6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CG3488 protein.
GN CG3488 OR ANON-23DA OR ANON-23D.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Chertoff E.H., Gelbart W.M.;
RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.Y.,
RA Brandon R.C., Rogers Y.-H.C., Blaese R.G., Champ M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin Z.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```



```
DR EMBL; BC000698; AAH00698.1; -.
DR EMBL; BC004253; AAH04253.1; -.
DR EMBL; BC008636; AAH08636.1; -.
DR EMBL; BC020982; AAH20982.1; -.
DR EMBL; M26326; AAAS9463.1; -.
DR EMBL; X12876; CAA31369.1; -.
DR EMBL; X12881; CAA31375.1; -.
DR EMBL; X12883; CAA31377.1; -.
DR EMBL; M26325; AAAS9462.1; -.
DR PIR; S05481; S05481.
DR GLYCOSUITE; P05783; -.
DR SWISS-2DPAGE; P05783; HUMAN.
DR PMMA-2DPAGE; P05783; -.
DR SIENA-2DPAGE; P05783; -.
DR GENE; HGNC:6430; KRT18.
DR MIM; 148070; -.
DR GO; GO:0005882; C:intermediate filament; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
DR K W Intermediate filament; Coiled coil; Keratin; Glycoprotein;
FT INIT_MET 0 0
FT MOD_RES 0 1
FT DOMAIN 1 78 ACETYLATION.
FT DOMAIN 79 386 HEAD.
FT DOMAIN 387 429 ROD.
FT DOMAIN 79 114 TAIL.
FT DOMAIN 115 131 COIL 1A.
FT DOMAIN 132 223 LINKER 1.
FT DOMAIN 224 247 COIL 1B.
FT DOMAIN 248 386 LINKER 12.
FT SITE 270 COIL 2.
FT SITE 330 STUTTER.
FT CARBOHYD 29 29 O-LINKED (GLCNAC).
FT CARBOHYD 30 30 /FTID-CAR_000175.
FT CARBOHYD 48 48 O-LINKED (GLCNAC).
FT CARBOHYD 48 48 /FTID-CAR_000193.
FT VARIANT 127 127 H -> L (IN CC; INTERFERS WITH THE ABILITY
FT TO FORM NORMAL FILAMENTS).
FT CONFLICT 167 167 /FTID=VAR_003852.
FT CONFLICT 201 201 Y -> H (IN REF. 2; AAH00698).
FT CONFLICT 245 245 E -> Q (IN REF. 4).
FT CONFLICT 308 308 A -> S (IN REF. 4).
FT CONFLICT 311 311 D -> R (IN REF. 4).
FT CONFLICT 311 311 S -> R (IN REF. 4).
SQ SEQUENCE 429 AA; 47926 MW; 717E9D6A3933F8 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
DB 390 NLGDALD 396
|||||

RESULT 36
SYTL_CABEL STANDARD; PRT; 441 AA.
AC F34693;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Synaptotagmin I.
GN SNT-1 OR F31E8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
```

```
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=93313960; PubMed=8391930;
RA Nonet M.L., Grundahl K., Meyer B.J., Rand J.B.;
RT "Synaptic function is impaired but not eliminated in C. elegans
RL mutants lacking synaptotagmin.";
RL Cell 73:1291-1305(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY HAVE A REGULATORY ROLE IN THE MEMBRANE INTERACTIONS
CC DURING TRAFFICKING OF SYNAPTIC VESICLES AT THE ACTIVE ZONE OF THE
CC SYNAPSE. IT BINDS ACIDIC PHOSPHOLIPIDS WITH A SPECIFICITY THAT
CC REQUIRES THE PRESENCE OF BOTH AN ACIDIC HEAD GROUP AND A DIACYL
CC BACKBONE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SYNAPTIC VESICLES AND VESICLE-LIKE
CC STRUCTURES.
CC -!- TISSUE SPECIFICITY: LOCALIZED TO REGIONS KNOWN TO BE RICH IN
CC SYNAPSES AND APPEARS TO BE ASSOCIATED WITH SYNAPTIC VESICLES.
CC ALSO FOUND IN SOME NONNEURONAL SECRETORY STRUCTURES.
CC -!- DISEASE: MUTANTS EXHIBIT SEVERE BEHAVIORAL ABNORMALITIES THAT ARE
CC CHARACTERISTIC OF DEFICIENCIES IN SYNAPTIC FUNCTION, INCLUDING
CC SEVERE LOCOMOTION, FEEDING, AND DEFECATION DEFECTS.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: BELONGS TO THE SYNAPTOTAGMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; LI5302; AAA28145.1; -.
DR EMBL; U55856; AAA98023.1; -.
DR PIR; A40707; A40707.
DR HSPP; P21707; IRSY.
DR WormPep; F31E8.2; CE02711.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002149; IRI.
DR Pfam; PF00168; C2; 2.
DR PRINTS; PR00360; C2DOMAIN.
DR PRINTS; PR00399; SYNAPTOTAGMIN.
DR SMART; SM00239; C2; 2.
DR PROSITE; PS00499; C2_DOMAIN_1; 2.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
KW Transmembrane; Repeat; Synapse; Glycoprotein.
FT DOMAIN 1 69 VESICULAR (POTENTIAL).
FT TRANSMEM 70 96 POTENTIAL.
FT DOMAIN 97 441 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 175 262 C2 DOMAIN 1.
FT DOMAIN 306 397 C2 DOMAIN 2.
SQ SEQUENCE 441 AA; 49904 MW; F8D174337EB472DB CRC64;

Query Match 0.9%; Score 7; DB 1; Length 441;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 TSEAVKQ 418
DB 32 TSEAVKQ 38
|||||

RESULT 37
CD4_CANFA STANDARD; PRT; 463 AA.
ID_CD4_CANFA
```

AC P33705;
 DT 01-FEB-1994 (Rel. 28, Last created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
 DE T4/Leu-3).
 GN CD4.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE OF 13-463 FROM N.A.
 RC STRAIN=Beagle; TISSUE=Thymus;
 RC MEDLINE=94378217; PubMed=8091416;
 RA Gorman S.D., Frewin M.R., Cobbold S.P., Waldmann H.;
 RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
 alpha antigens.";
 RL Tissue Antigens 43:184-188(1994).
 CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
 CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION.
 CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MACROPHAGES AND A SUBSET OF
 CC T LYMPHOCYTES.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L06130; AAB02295.1; -;
 DR EMBL; X68565; -; NOT_ANNOTATED_CDS.
 DR HSP; P01730; 1WBR.
 DR GO; GO:0042101; C:T-cell receptor complex; ISS.
 DR GO; GO:0015026; F:coreceptor activity; ISS.
 DR GO; GO:0042289; F:MHC class II protein binding activity; ISS.
 DR GO; GO:0006955; P:immune response; ISS.
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
 DR GO; GO:0030217; P:T-cell differentiation; ISS.
 DR GO; GO:0045058; P:T-cell selection; ISS.
 DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
 DR InterPro; IPR000973; CD4 TCAG.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig.MHC.
 DR Pfam; PF00047; Ig. 3.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV. 1.
 DR PROSITE; PS50835; IG LIKE; 1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
 FT SIGNAL 1 24
 FT CHAIN 25 463 T-CELL SURFACE GLYCOPROTEIN CD4.
 FT DOMAIN 25 401 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 402 423 POTENTIAL.
 FT DOMAIN 424 463 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 124 IG-LIKE V-TYPE.
 FT DOMAIN 125 211 IG-LIKE C2-TYPE 1.
 FT DOMAIN 212 321 IG-LIKE C2-TYPE 2.
 FT DOMAIN 322 378 IG-LIKE C2-TYPE 3.

FT DISULFID 41 109 BY SIMILARITY.
 FT DISULFID 332 374 BY SIMILARITY.
 FT LIPID 424 424 PALMITATE (BY SIMILARITY).
 FT LIPID 427 427 PALMITATE (BY SIMILARITY).
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 389 389 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 463 AA; 51639 MW; 95805170CB4A833 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 463;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 365 KAGDAVE 371
 Db |||||
 Db 32 KAGDAVE 38
 RESULT 38
 SG3_MOUSE
 ID SG3_MOUSE STANDARD; PRT; 471 AA.
 AC P47867;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Secretogranin III precursor (SgIII).
 GN SG3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95001263; PubMed=7917832;
 RA Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
 RA Sutcliffe J.G.;
 RT "Primary structure of mouse secretogranin III and its absence from
 RT deficient mice.";
 R J. Mol. Neurosci. 4:225-233(1993).
 CC -1- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
 CC secretory granules.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U02982; AAA56636.1; -;
 DR MGD; MGI:103032; Scg3.
 KW Signal; Cleavage on pair of basic residues.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 471 SECRETOTRANIN III.
 SQ SEQUENCE 471 AA; 53326 MW; FC1E9C381AFCA564 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 471;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 160 DGKIPEK 166
 Db |||||
 Db 226 DGKIPEK 232
 RESULT 39
 SG3_RAT
 ID SG3_RAT STANDARD; PRT; 471 AA.

AC P47868;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Secretogranin III precursor (SgIII) (1B1075).
GN SCG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95001263; PubMed=7917832;
RA Dopazo A., Lovenberg T.W., Danielson P.E., Ottiger H.-P.,
RA Sutcliffe J.G.;
RT "Primary structure of mouse secretogranin III and its absence from
RT deficient mice."
RL J. Mol. Neurosci. 4:225-233(1993).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=90376160; PubMed=2204688;
RA Ottiger H.-P., Battenberg E.F., Tsou A.-P., Bloom F.E.,
RA Sutcliffe J.G.;
RT "1B1075: a brain- and pituitary-specific mRNA that encodes a novel
RT chromogranin / secretogranin-like component of intracellular
RT vesicles."
RL J. Neurosci. 10:3135-3147(1990).
CC -1- SUBCELLULAR LOCATION: Secreted. Neuroendocrine and endocrine
CC -1- TISSUE SPECIFICITY: BRAIN- AND PITUITARY-SPECIFIC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U02983; AAA56637.1; -
KW Signal; Cleavage on pair of basic residues.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 471 SECRETGRANIN III.
SQ SEQUENCE 471 AA; 53183 MW; 977B3F8885F33223 CRC64;
Query Match 0.9%; Score 7; DB 1; Length 471;
Best local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 160 DGKPEK 166
Db 226 DGKPEK 232
RESULT 40
Y213 MYCGE
ID Y213 MYCGE STANDARD; PRT; 471 AA.
AC P47455;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG213.
GN MG213.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Fritchman J.L., Weidman J.P., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudak D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -1- SIMILARITY: SOME, TO B.SUBTILIS YPUG.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U39701; AAC71432.1; -
DR PIR; E64223; E64223.
DR TIGR; MG213; -
DR InterPro; IPR003768; DUF173.
DR Pfam; PF02616; DUF173; 1.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 422 427 POLY-GLN.
SQ SEQUENCE 471 AA; 54829 MW; 9531DEC11D790FE2 CRC64;
Query Match 0.9%; Score 7; DB 1; Length 471;
Best local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 724 KADNLLN 730
Db 217 KADNLLN 223
Search completed: November 14, 2003, 11:07:16
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:05:49 ; Search time 21 Seconds

(without alignments)
3471.231 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 758

Sequence: 1 MAQTTLKPVLISILLINTPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 9616682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

PIR 76:*
1: Piri:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	175	23.1	764	D81893	probable outer-mem
2	9	1.2	205	A72479	hypothetical prote
3	9	1.2	323	T05478	peroxidase (EC 1.1
4	9	1.2	1980	S54307	myosin heavy chain
5	9	1.2	2022	A59256	myosin-IxB [simila
6	8	1.1	99	AB1484	hypothetical prote
7	8	1.1	99	AB1123	hypothetical prote
8	8	1.1	196	AE2883	nitroreductase [l
9	8	1.1	196	C97659	probable enzyme (A
10	8	1.1	316	JUNBSLC	beta-lactamase (EC
11	8	1.1	393	JU0334	formate dehydrogen
12	8	1.1	399	T19180	hypothetical prote
13	8	1.1	400	JC7815	formate dehydrogen
14	8	1.1	401	C95233	probable NAD-depen
15	8	1.1	544	S76859	hypothetical prote
16	8	1.1	892	T40040	Grpase-activator p
17	8	1.1	1009	S61174	hypothetical prote
18	7	0.9	54	S34567	gene L protein - h
19	7	0.9	76	AB1783	hypothetical prote
20	7	0.9	100	E89460	conserved hypotet
21	7	0.9	108	A72224	hypothetical prote
22	7	0.9	117	S01185	NADH2 dehydrogenas
23	7	0.9	117	G25797	NADH2 dehydrogenas
24	7	0.9	117	C87588	hypothetical prote
25	7	0.9	125	T05445	hypothetical prote
26	7	0.9	141	T27080	hypothetical prote
27	7	0.9	141	T27071	hypothetical prote
28	7	0.9	149	T35846	probable integral
29	7	0.9	153	DTECR	aspartate carbamoy

30	7	0.9	153	DTEBCT	aspartate carbamoy
31	7	0.9	153	E91281	aspartate carbamoy
32	7	0.9	153	F86122	aspartate carbamoy
33	7	0.9	153	AC1059	hypothetical prote
34	7	0.9	155	S76912	hypothetical prote
35	7	0.9	166	T07089	dehydrin - soybean
36	7	0.9	171	AF2864	acetyltransferase
37	7	0.9	171	E97641	hypothetical prote
38	7	0.9	173	C70228	conserved hypotet
39	7	0.9	185	A87609	hypothetical prote
40	7	0.9	200	A41740	H+-transporting tw
41	7	0.9	201	F72285	transcription regu
42	7	0.9	202	F72739	hypothetical prote
43	7	0.9	212	B84337	deoxyribose-phosph
44	7	0.9	215	T33566	hypothetical prote
45	7	0.9	225	D64314	conserved hypotet
46	7	0.9	230	I37095	gene 2.19 protein
47	7	0.9	234	A90402	flagella-related p
48	7	0.9	241	G89925	menaquinone biosyn
49	7	0.9	244	H69844	diadenosine tetrap
50	7	0.9	249	D87533	hypothetical prote
51	7	0.9	249	S09868	hypothetical prote
52	7	0.9	249	E72646	probable molybdopt
53	7	0.9	250	B69038	conserved hypotet
54	7	0.9	251	AG0132	probable hydroxyac
55	7	0.9	251	T32200	hypothetical prote
56	7	0.9	263	A72349	conserved hypotet
57	7	0.9	265	T50569	probable membrane
58	7	0.9	278	T33129	hypothetical prote
59	7	0.9	280	D69513	conserved hypotet
60	7	0.9	285	A84386	hypothetical prote
61	7	0.9	294	T26946	hypothetical prote
62	7	0.9	295	B82674	conserved hypotet
63	7	0.9	303	B88241	protein T22C8.1 [l
64	7	0.9	303	T25114	hypothetical prote
65	7	0.9	310	C83909	oligopeptide ABC t
66	7	0.9	312	F87126	probable short cha
67	7	0.9	312	G82105	flagellar protein
68	7	0.9	315	A84598	probable triosepho
69	7	0.9	316	B81321	probable cation tr
70	7	0.9	320	T33662	hypothetical prote
71	7	0.9	325	TJ7553	brain mitochondria
72	7	0.9	327	C27483	hypothetical prote
73	7	0.9	333	JS0590	endo-1,4-beta-xyla
74	7	0.9	335	T50601	endo-1,4-beta-xyla
75	7	0.9	338	F84262	hypothetical prote
76	7	0.9	343	T26594	hypothetical prote
77	7	0.9	349	G69351	N2, N2-dimethylguan
78	7	0.9	353	T48310	hypothetical prote
79	7	0.9	368	A82093	membrane-bound lyl
80	7	0.9	369	JQ2278	hydroxymethylbilan
81	7	0.9	388	T37995	alanine racemase (
82	7	0.9	389	A38302	pepsin (EC 3.4.23.
83	7	0.9	391	AD2501	hypothetical prote
84	7	0.9	394	B70206	hypothetical prote
85	7	0.9	401	T18936	phosphoprotein pho
86	7	0.9	403	T87270	general secretion
87	7	0.9	406	H70364	iron-sulfur cofact
88	7	0.9	414	A12823	aminotransferase (
89	7	0.9	414	H97601	hypothetical prote
90	7	0.9	415	C70552	hypothetical prote
91	7	0.9	430	S05481	keratin 18, type I
92	7	0.9	432	S30193	T-cell surface gly
93	7	0.9	434	D72353	lipopolysaccharide
94	7	0.9	437	T34211	hypothetical prote
95	7	0.9	441	A40707	synaptotagmin - Ca
96	7	0.9	442	C96688	unknown protein, 2
97	7	0.9	443	B86180	protein TlG11.3 [i
98	7	0.9	444	JC4348	virulence-mediati
99	7	0.9	448	T01814	hypothetical prote
100	7	0.9	468	B81924	probable two-compo
101	7	0.9	468	G81180	sensor histidine k
102	7	0.9	468	T22287	hypothetical prote

103	7	0.9	469	2	D82976	hypothetical prote	176	7	0.9	2142	2	D86303	F17F16.1 protein -
104	7	0.9	469	2	T48783	related to SGT1 pr	177	7	0.9	3263	2	D82410	hypothetical prote
105	7	0.9	469	2	E86421	hypothetical prote	178	7	0.9	3433	1	S28381	utrophin - human
106	7	0.9	470	2	AF3435	tldD protein limpo	179	7	0.9	4302	2	A38971	polycystic kidney
107	7	0.9	471	2	E64223	hypothetical prote	180	7	0.9	4485	2	T08044	dynein gamma heavy
108	7	0.9	488	2	A81606	glycine dehydrogen	181	6	0.8	20	2	A49142	hyaluronidase (EC
109	7	0.9	488	2	AF1243	glycine dehydrogen	182	6	0.8	22	2	T09010	N4-(beta-N-acetyl)g
110	7	0.9	490	2	A83433	mitochondrial proc	183	6	0.8	27	2	B43853	probable phosphogl
111	7	0.9	492	2	T02458	hypothetical prote	184	6	0.8	33	2	T08018	ycf12 protein - Ch
112	7	0.9	497	2	T48676	proline uptake pro	185	6	0.8	50	2	AG2783	hypothetical prote
113	7	0.9	498	2	A87374	hypothetical prote	186	6	0.8	50	2	G97562	hypothetical prote
114	7	0.9	509	2	D86794	unknown Protein F1	187	6	0.8	51	2	PN0479	cell division cycl
115	7	0.9	512	2	E89978	high affinity prol	188	6	0.8	57	2	H95181	hypothetical prote
116	7	0.9	513	2	T37180	probable membrane	189	6	0.8	57	2	C98049	hypothetical prote
117	7	0.9	524	2	D70861	probable monooxygen	190	6	0.8	59	2	E84011	hypothetical prote
118	7	0.9	533	2	T13607	hypothetical prote	191	6	0.8	60	2	A10343	probable regulator
119	7	0.9	533	2	A37180	chromogranin/secret	192	6	0.8	62	2	G64561	hypothetical prote
120	7	0.9	537	2	B84772	probable Bnad prot	193	6	0.8	62	2	D89907	hypothetical prote
121	7	0.9	537	2	G31277	guinate transport	194	6	0.8	64	2	F86693	hypothetical prote
122	7	0.9	539	2	S58287	J-domain protein D	195	6	0.8	66	2	B99970	hypothetical prote
123	7	0.9	545	2	S46151	probable purine nu	196	6	0.8	70	2	T44130	hypothetical prote
124	7	0.9	555	2	H83043	hypothetical prote	197	6	0.8	71	2	AF2940	hypothetical prote
125	7	0.9	558	2	D70449	conserved hypothet	198	6	0.8	71	2	C98342	hypothetical prote
126	7	0.9	569	2	T50711	urease (EC 3.5.1.5	199	6	0.8	72	2	A12928	conserved hypothet
127	7	0.9	594	2	E96667	unknown protein, 6	200	6	0.8	72	2	E98353	hypothetical prote
128	7	0.9	601	2	B75136	hypothetical prote	201	6	0.8	73	2	B83338	hypothetical prote
129	7	0.9	608	2	A35021	myosin-light-chain	202	6	0.8	73	2	F83266	hypothetical prote
130	7	0.9	623	2	H87127	conserved hypothet	203	6	0.8	75	2	A13219	hypothetical prote
131	7	0.9	628	2	A87596	hypothetical prote	204	6	0.8	82	2	F95086	conserved hypothet
132	7	0.9	642	2	E87644	sensory box histid	205	6	0.8	82	2	A37954	conserved hypothet
133	7	0.9	645	2	D85680	unknown protein en	206	6	0.8	85	2	S14026	hypothetical prote
134	7	0.9	645	2	D85844	major head protein	207	6	0.8	85	2	C82589	hypothetical prote
135	7	0.9	645	2	A85629	major head protein	208	6	0.8	87	2	T09816	26S proteinase reg
136	7	0.9	645	2	A90853	major head protein	209	6	0.8	87	2	T14135	NADH2 dehydrogenas
137	7	0.9	645	2	G90821	major head protein	210	6	0.8	89	2	T09819	26S proteinase reg
138	7	0.9	645	2	D90910	major head protein	211	6	0.8	89	2	C72286	transcription regu
139	7	0.9	645	2	D90875	major head protein	212	6	0.8	92	2	PQ6632	coat protein - Rem
140	7	0.9	648	2	S59723	transcription acti	213	6	0.8	94	2	D82498	conserved hypothet
141	7	0.9	658	2	C75187	anaerobic ribonucl	214	6	0.8	95	2	A03858	hypothetical prote
142	7	0.9	676	2	S70395	env polyprotein -	215	6	0.8	100	2	T38899	hypothetical prote
143	7	0.9	720	2	S69890	mitogen inducible	216	6	0.8	100	2	H81310	probable periplasm
144	7	0.9	728	2	T27884	hypothetical prote	217	6	0.8	100	2	AH2220	hypothetical prote
145	7	0.9	737	2	AD2037	hypothetical prote	218	6	0.8	100	2	A10092	probable membrane
146	7	0.9	741	2	T13791	NADH2 dehydrogenas	219	6	0.8	101	1	W7ML51	E7 protein - human
147	7	0.9	745	2	C88618	protein ZK520.4 [i	220	6	0.8	101	2	JX0333	ribonuclease (EC 3
148	7	0.9	764	2	H83055	probable outer mem	221	6	0.8	101	2	A43262	hypothetical prote
149	7	0.9	765	2	AF2200	phytochrome A, two	222	6	0.8	101	2	E83252	hypothetical prote
150	7	0.9	765	2	T15447	hypothetical prote	223	6	0.8	103	2	I79265	traa protein - Esc
151	7	0.9	770	2	T26783	hypothetical prote	224	6	0.8	104	1	ASLJ5Y	vpr protein - huma
152	7	0.9	782	2	E90427	hypothetical prote	225	6	0.8	104	2	T03316	gene 19 protein -
153	7	0.9	802	2	AF1227	phenylalanyl-tRNA	226	6	0.8	104	2	D95232	hypothetical prote
154	7	0.9	802	2	AH1580	phenylalanyl-tRNA	227	6	0.8	104	2	F98096	hypothetical prote
155	7	0.9	818	2	F97772	RNA-directed RNA p	228	6	0.8	107	2	B71069	hypothetical prote
156	7	0.9	844	1	RRXSSP	hypothetical prote	229	6	0.8	108	2	B71074	hypothetical prote
157	7	0.9	847	2	T04772	alanyl-tRNA synthe	230	6	0.8	109	2	F89886	hypothetical prote
158	7	0.9	881	2	AG7107	transcription regu	231	6	0.8	110	2	C97137	hypothetical prote
159	7	0.9	892	2	AG1661	CTF4 protein - yea	232	6	0.8	112	2	C72679	hypothetical prote
160	7	0.9	927	2	A45039	penicillin-binding	233	6	0.8	112	2	B84345	hypothetical prote
161	7	0.9	966	2	E84053	rho-type GTPase-ac	234	6	0.8	113	2	B83575	hypothetical prote
162	7	0.9	1007	2	A48535	probable RND efflu	235	6	0.8	115	2	A84676	60S acidic ribosom
163	7	0.9	1018	2	E83099	hypothetical prote	236	6	0.8	115	2	G72466	hypothetical prote
164	7	0.9	1044	2	T37568	TonB-dependent rec	237	6	0.8	116	2	S14211	NADH2 dehydrogenas
165	7	0.9	1046	2	H87318	hypothetical prote	238	6	0.8	116	2	S59099	ribosomal protein
166	7	0.9	1052	1	B49219	toxin iii - Actino	239	6	0.8	116	2	C69071	hypothetical prote
167	7	0.9	1067	2	S33417	kinesin-like prote	240	6	0.8	117	2	C69053	ribosomal protein
168	7	0.9	1086	2	T33893	hypothetical prote	241	6	0.8	117	2	F82308	hypothetical prote
169	7	0.9	1186	2	AG1928	two-component hybr	242	6	0.8	118	2	JE0398	ribosomal protein
170	7	0.9	1252	2	H97178	ATP-dependent exon	243	6	0.8	118	2	S70089	korA protein - Amy
171	7	0.9	1366	2	T35985	probable large Pro	244	6	0.8	119	1	YQECB2	fimbrial protein p
172	7	0.9	1400	2	T52359	hypothetical prote	245	6	0.8	119	1	YQECR1	fimbrial protein p
173	7	0.9	1786	1	H71527	probable excinucle	246	6	0.8	119	2	S03302	T-cell receptor ga
174	7	0.9	1832	2	T31113	mucin-like glycopr	247	6	0.8	119	2	I40630	prepilin - Escheri
175	7	0.9	2089	1	A48757	acetyl-CoA carboxy	248	6	0.8	119	2	C70140	hypothetical prote

249	6	0.8	120	1	YQECR9	fimbrial protein p	322	6	0.8	149	2	F83964	hypothetical prote
250	6	0.8	120	2	T12918	hypothetical prote	323	6	0.8	150	2	C97826	hypothetical prote
251	6	0.8	121	1	RIAFSI	somatostatin I pre	324	6	0.8	150	2	E85514	unknown protein fr
252	6	0.8	121	1	YQECF	fimbrial protein p	325	6	0.8	150	2	H97158	hypothetical prote
253	6	0.8	121	1	YQECF	fimbrial protein p	326	6	0.8	151	2	T43478	hypothetical prote
254	6	0.8	121	1	B2AG58	virB2 protein prec	327	6	0.8	152	2	C81189	probable periplasm
255	6	0.8	121	2	AF3248	hypothetical prote	328	6	0.8	152	2	F72646	hypothetical prote
256	6	0.8	121	2	D72553	hypothetical prote	329	6	0.8	152	2	AI3271	acetyltransferase
257	6	0.8	121	2	D82604	hypothetical prote	330	6	0.8	152	2	D98270	hypothetical prote
258	6	0.8	122	2	C75317	hypothetical prote	331	6	0.8	153	2	S08501	3-dehydroquinat d
259	6	0.8	122	2	T01558	auxin-induced prot	332	6	0.8	153	2	AF2203	bacterioferritin c
260	6	0.8	122	2	S48325	probable membrane	333	6	0.8	153	2	AF2344	hypothetical prote
261	6	0.8	123	1	RHPCN	galanin precursor	334	6	0.8	153	2	T31654	hypothetical prote
262	6	0.8	124	2	A83505	hypothetical prote	335	6	0.8	154	1	XMECF1	flagellar protein
263	6	0.8	124	2	S75364	hypothetical prote	336	6	0.8	154	2	C90964	flagellar protein
264	6	0.8	124	2	A64733	probable membrane	337	6	0.8	154	2	C85812	flagellar biosynth
265	6	0.8	125	2	A96931	nitrogen regulator	338	6	0.8	154	2	E70902	probable riboflavi
266	6	0.8	126	2	S74493	thioredoxin M-2 -	339	6	0.8	154	2	E90794	urease accessory p
267	6	0.8	126	2	C33989	Ig heavy chain V-3	340	6	0.8	154	2	G85603	probable urease ac
268	6	0.8	128	2	A97248	nudix (Mnt) fami	341	6	0.8	155	1	WMBV7B	17K protein - barl
269	6	0.8	128	2	D70742	hypothetical prote	342	6	0.8	155	2	A82250	phosphotyrosine pr
270	6	0.8	129	2	F95865	probable ribulose-	343	6	0.8	155	2	S33509	interleukin-2 - Mo
271	6	0.8	129	2	S40332	Ig kappa chain - h	344	6	0.8	155	2	B30929	flagellar protein
272	6	0.8	129	2	C75328	hypothetical prote	345	6	0.8	155	2	AH0752	FluL protein (impo
273	6	0.8	130	2	A24574	hypothetical prote	346	6	0.8	155	2	A64801	yhey protein - Esc
274	6	0.8	130	2	D89817	T-cell receptor ga	347	6	0.8	155	2	A90716	hypothetical prote
275	6	0.8	130	2	C70971	hypothetical prote	348	6	0.8	155	2	D85566	hypothetical prote
276	6	0.8	130	2	C88102	protein W09G10.6 l	349	6	0.8	155	2	T34169	hypothetical prote
277	6	0.8	130	2	AC2314	hypothetical prote	350	6	0.8	156	2	AG0754	patch repair prote
278	6	0.8	131	2	H83156	hypothetical prote	351	6	0.8	157	2	E72332	thioredoxin peroxi
279	6	0.8	131	2	C83863	hypothetical prote	352	6	0.8	158	2	E86993	hypothetical prote
280	6	0.8	132	2	D72616	hypothetical prote	353	6	0.8	158	2	D36138	ureE protein - Kle
281	6	0.8	133	2	B82977	conserved hypothec	354	6	0.8	158	2	H83824	hypothetical prote
282	6	0.8	134	2	D96568	probable histone H	355	6	0.8	159	2	B84669	hypothetical prote
283	6	0.8	134	2	G82411	conserved hypothet	356	6	0.8	160	2	T24339	hypothetical prote
284	6	0.8	134	2	AG2552	hypothetical prote	357	6	0.8	160	2	H90985	hypothetical prote
285	6	0.8	134	2	D95296	hypothetical prote	358	6	0.8	160	2	C85831	GBP-mannose mannos
286	6	0.8	136	2	S12847	T-cell receptor ga	359	6	0.8	160	2	E55239	GNP-mannose mannos
287	6	0.8	136	2	T47615	histone H2A.F/Z -	360	6	0.8	160	2	A81075	Colanic acid biosy
288	6	0.8	136	2	E36790	hypothetical prote	361	6	0.8	161	2	T06785	FKBP-type peptidyl
289	6	0.8	136	2	F72353	flagellar protein	362	6	0.8	161	2	QJ2160	nonsymbiotic hemog
290	6	0.8	136	2	S76416	hypothetical prote	363	6	0.8	161	2	QJ2146	coat protein - tom
291	6	0.8	137	2	E69359	hypothetical prote	364	6	0.8	161	2	F64038	coat protein - tom
292	6	0.8	138	2	A37330	venom allergen II	365	6	0.8	161	2	C65170	hypothetical prote
293	6	0.8	139	2	B69386	conserved hypothet	366	6	0.8	162	2	D82983	pcs system, arbuti
294	6	0.8	139	2	A83579	conserved hypothet	367	6	0.8	162	2	T13522	leucine-responsive
295	6	0.8	139	2	H90298	hypothetical prote	368	6	0.8	162	2	D98020	hypothetical prote
296	6	0.8	139	2	G97240	amino acid transpo	369	6	0.8	162	2	B70144	conserved hypothet
297	6	0.8	140	2	T04904	hypothetical prote	370	6	0.8	163	2	B70854	transcription fact
298	6	0.8	141	2	JT0624	hemoglobin alpha 1	371	6	0.8	163	2	T10220	hypothetical prote
299	6	0.8	141	2	F69948	phage-related prot	372	6	0.8	164	2	B82888	hypothetical prote
300	6	0.8	141	2	AE3275	hypothetical cytos	373	6	0.8	164	2	S11846	single-stranded DN
301	6	0.8	143	2	S13238	tail fiber assembl	374	6	0.8	164	2	S11847	rab16B protein - r
302	6	0.8	145	1	ZWECAP	muca protein - Esc	375	6	0.8	164	2	S77233	rab16C protein - r
303	6	0.8	145	2	F83834	hypothetical prote	376	6	0.8	164	2	S77233	riboflavin synthas
304	6	0.8	145	2	C72668	hypothetical prote	377	6	0.8	164	2	I40526	hypothetical prote
305	6	0.8	146	2	S06775	hypothetical prote	378	6	0.8	164	2	H89784	conserved hypothet
306	6	0.8	146	2	C83805	muca protein - Sal	379	6	0.8	165	2	T15780	hypothetical prote
307	6	0.8	146	2	B64472	hypothetical prote	380	6	0.8	165	2	S34270	hypothetical prote
308	6	0.8	146	2	T10511	hypothetical prote	381	6	0.8	165	2	S34269	fibrinogen-binding
309	6	0.8	146	2	AD0929	conserved hypothet	382	6	0.8	166	2	G75449	fibrinogen-binding
310	6	0.8	146	2	G65008	hypothetical prote	383	6	0.8	167	2	T36325	MutT/nudix family
311	6	0.8	147	2	A29910	myosin calcium-bin	384	6	0.8	167	2	D55224	hypothetical prote
312	6	0.8	147	2	C64514	hypothetical prote	385	6	0.8	167	2	S34363	hypothetical prote
313	6	0.8	147	2	G75629	conserved hypothet	386	6	0.8	168	2	H69940	phosphotransferase
314	6	0.8	147	2	H70630	hypothetical prote	387	6	0.8	168	2	S72898	hypothetical prote
315	6	0.8	147	2	B90664	hypothetical prote	388	6	0.8	168	2	T45305	hypothetical prote
316	6	0.8	148	2	T52550	probable transcrip	389	6	0.8	169	2	E26395	hypothetical prote
317	6	0.8	148	2	E95886	probable sugar-pho	390	6	0.8	170	2	H82746	T-cell receptor be
318	6	0.8	148	2	H64004	hypothetical prote	391	6	0.8	171	2	S75677	polypeptide deform
319	6	0.8	148	2	F95994	hypothetical prote	392	6	0.8	171	2	A11060	hypothetical prote
320	6	0.8	149	2	C97180	hypothetical prote	393	6	0.8	172	2	T27505	probable acetyltra
321	6	0.8	149	2	D82178	uncharacterized pr	394	6	0.8	173	2	H95961	hypothetical prote
						hypothetical prote				174	1	A31096	fimbrial protein p

395	6	0.8	174	2	AF3297	biopolymer transpo	468	6	0.8	201	2	H82935	recombinational DN
396	6	0.8	174	2	T39690	probable actin pol	469	6	0.8	201	2	C88473	protein F40Hc.3 [i
397	6	0.8	175	2	F37802	criz protein - Erv	470	6	0.8	201	2	R83319	hypothetical prote
398	6	0.8	175	2	AC3014	MutT/nudix family	471	6	0.8	202	1	MNIV47	nonstructural prot
399	6	0.8	175	2	G84383	hypothetical prote	472	6	0.8	202	2	E75209	hypothetical prote
400	6	0.8	176	2	F82478	hypothetical prote	473	6	0.8	202	2	T43982	glycoprotein limpo
401	6	0.8	176	2	B70445	heat shock protein	474	6	0.8	202	2	T44168	hypothetical prote
402	6	0.8	178	2	D69467	probable orotate p	475	6	0.8	202	2	G86302	FL7Fl6.6 protein -
403	6	0.8	178	2	T20317	hypothetical prote	476	6	0.8	203	2	S51495	GTP-binding protei
404	6	0.8	179	2	G89817	hypothetical prote	477	6	0.8	203	2	F75554	MutT/nudix family
405	6	0.8	179	2	AT2056	polypeptide deform	478	6	0.8	204	2	H96027	probable protocate
406	6	0.8	179	2	T40442	hypothetical highl	479	6	0.8	204	2	D64334	hypothetical prote
407	6	0.8	180	2	S32917	hypothetical prote	480	6	0.8	204	2	C97866	hypothetical prote
408	6	0.8	180	2	H81181	bacteriocin resist	481	6	0.8	205	2	AB3600	hypothetical prote
409	6	0.8	181	2	T24914	hypothetical prote	482	6	0.8	205	2	AE3432	hypothetical prote
410	6	0.8	181	2	T33410	hypothetical prote	483	6	0.8	206	2	AG3114	2-haloacid dehalog
411	6	0.8	181	2	AG3915	hypothetical prote	484	6	0.8	206	2	F98172	protocatechuate 3,
412	6	0.8	181	2	F89927	conserved hypothet	485	6	0.8	207	2	S54128	hypothetical 21.4K
413	6	0.8	182	2	F81366	adenine phosphorib	486	6	0.8	207	2	AB3016	hypothetical prote
414	6	0.8	182	2	C87097	conserved hypothet	487	6	0.8	207	2	C90534	general secretion
415	6	0.8	182	2	F57899	hypothetical prote	488	6	0.8	208	2	AB1877	glutathione transf
416	6	0.8	182	2	F37985	hypothetical prote	489	6	0.8	208	2	S03615	ribosomal protein
417	6	0.8	183	1	TUBP84	tail fiber assembl	490	6	0.8	208	2	I64209	hypothetical prote
418	6	0.8	183	1	S13240	hypothetical prote	491	6	0.8	208	2	F87633	protocatechuate 3,
419	6	0.8	183	2	A89879	hypothetical prote	492	6	0.8	209	1	D35119	hypothetical prote
420	6	0.8	183	2	T07241	hypothetical prote	493	6	0.8	210	2	C75380	uracil phosphoribo
421	6	0.8	184	2	A61628	early gland protei	494	6	0.8	210	2	F82238	phosphoribosyl-AMP
422	6	0.8	184	2	B30592	hypothetical prote	495	6	0.8	210	2	F71308	probable 2-dehydro
423	6	0.8	185	2	C83644	conserved hypothet	496	6	0.8	210	2	T33697	hypothetical prote
424	6	0.8	187	2	H64995	hypothetical prote	497	6	0.8	210	2	C70528	hypothetical prote
425	6	0.8	187	2	B31021	hypothetical prote	498	6	0.8	210	2	A87331	hypothetical prote
426	6	0.8	187	2	A85865	hypothetical prote	499	6	0.8	210	2	T26154	hypothetical prote
427	6	0.8	187	2	A95896	conserved hypothet	500	6	0.8	210	2	A72710	hypothetical prote
428	6	0.8	188	2	AH0588	probable DNA recom	501	6	0.8	211	2	S73791	hypothetical prote
429	6	0.8	188	2	A84214	adenine phosphorib	502	6	0.8	211	2	T09316	glycoprotein - hum
430	6	0.8	189	2	A25556	hypothetical 21.2K	503	6	0.8	211	2	D97775	hypothetical prote
431	6	0.8	189	2	F96593	hypothetical prote	504	6	0.8	211	2	C96539	hypothetical prote
432	6	0.8	189	2	AG1947	hypothetical prote	505	6	0.8	212	1	F65054	L-fucose-phospha
433	6	0.8	190	2	S70699	X4-transporting AT	506	6	0.8	212	1	B64166	cytochrome c bioge
434	6	0.8	190	2	A82305	6,7-dimethyl-8-rib	507	6	0.8	212	2	AG0854	probable sugar ald
435	6	0.8	191	2	T31903	hypothetical prote	508	6	0.8	212	2	A71702	hypothetical prote
436	6	0.8	191	2	T21031	hypothetical prote	509	6	0.8	212	2	B69493	hypothetical prote
437	6	0.8	192	2	T17189	conserved hypothet	510	6	0.8	214	2	B95898	probable sensory t
438	6	0.8	192	2	A82277	tellurite resistan	511	6	0.8	214	2	T33352	hypothetical prote
439	6	0.8	192	2	G87452	hypothetical prote	512	6	0.8	214	2	T22487	hypothetical prote
440	6	0.8	192	2	A86398	protein T7N9.22 [i	513	6	0.8	215	2	A49677	endoplasmic reticu
441	6	0.8	192	2	S67201	hypothetical prote	514	6	0.8	215	2	I51377	endymidin precurs
442	6	0.8	193	2	G82126	Maf/Ycep/YndE fami	515	6	0.8	215	2	G70551	hypothetical prote
443	6	0.8	193	2	T96581	hypothetical prote	516	6	0.8	215	2	C86962	probable oligoribo
444	6	0.8	194	2	I57523	HSF90 - mouse (fra	517	6	0.8	215	2	A80793	conserved hypothet
445	6	0.8	194	2	D75428	hypothetical prote	518	6	0.8	216	2	JI0090	ependymin precurs
446	6	0.8	194	2	T19641	hypothetical prote	519	6	0.8	216	2	A32636	ependymin II precu
447	6	0.8	194	2	T19642	hypothetical prote	520	6	0.8	216	2	S64240	probable membrane
448	6	0.8	195	2	AD2243	hypothetical prote	521	6	0.8	216	2	H84241	hypothetical prote
449	6	0.8	195	2	T19643	hypothetical prote	522	6	0.8	216	2	B87587	hypothetical prote
450	6	0.8	196	2	AD0385	4-methyl-5(B-hydro	523	6	0.8	216	2	AH0685	probable lipoprote
451	6	0.8	196	2	G72063	15 kDa Cysteine-ri	524	6	0.8	216	2	B87674	conserved hypothet
452	6	0.8	196	2	H86559	15 kDa Cysteine-ri	525	6	0.8	217	1	JS0630	formate dehydrogen
453	6	0.8	196	2	F90785	probable enzyme [i	526	6	0.8	217	1	MNIV61	nonstructural prot
454	6	0.8	196	2	D85645	probable enzyme z1	527	6	0.8	217	2	A85729	formate dehydrogen
455	6	0.8	196	2	F64842	hypothetical prote	528	6	0.8	217	2	H90888	formate dehydrogen
456	6	0.8	197	2	S46928	phytochrome - Char	529	6	0.8	217	2	F82788	thymidylate kinase
457	6	0.8	197	2	A81912	probable GTP cyclo	530	6	0.8	217	2	F70645	hypothetical prote
458	6	0.8	197	2	AG1104	GTP cyclohydrolase	531	6	0.8	217	2	T42053	redz protein - Str
459	6	0.8	197	2	A82556	c-type cytochrome	532	6	0.8	217	2	T42053	two component resp
460	6	0.8	197	2	S35252	proline-rich prote	533	6	0.8	217	2	AF3050	probable two-compo
461	6	0.8	197	2	T43066	hypothetical prote	534	6	0.8	217	2	A43820	ependymin precurs
462	6	0.8	198	2	H69487	SSU ribosomal prot	535	6	0.8	217	2	H70631	hypothetical prote
463	6	0.8	198	2	E81391	probable transcrip	536	6	0.8	217	2	AF1871	hypothetical prote
464	6	0.8	199	2	T08902	manganese-binding	537	6	0.8	217	2	T46468	hypothetical prote
465	6	0.8	199	2	T36622	hypothetical prote	538	6	0.8	218	2	F83857	phosphoribosyl ant
466	6	0.8	200	2	H84715	probable phytocyan	539	6	0.8	218	2	C97236	ABC transporter, A
467	6	0.8	200	2	B81322	probable integral	540	6	0.8	218	2	S73675	hypothetical prote

541	6	0.8	218	2	T25229	hypothetical prote	614	6	0.8	237	1	MNIV77	nonstructural prot
542	6	0.8	218	2	T03383	zinc-induced prote	615	6	0.8	237	1	MNIVC1	nonstructural prote
543	6	0.8	219	2	D71001	hypothetical prote	616	6	0.8	237	1	F84687	hypothetical prote
544	6	0.8	219	2	T02112	probable bZIP tran	617	6	0.8	237	4	S58729	probable ribosomal
545	6	0.8	219	2	E83399	hypothetical prote	618	6	0.8	238	2	H72778	hypothetical prote
545	6	0.8	219	2	AC2135	hypothetical prote	619	6	0.8	238	2	G83528	hypothetical prote
547	6	0.8	220	2	I50588	fibroblast growth	620	6	0.8	239	2	E64699	hypothetical prote
548	6	0.8	220	2	T39548	hypothetical prote	621	6	0.8	239	2	T49360	related to a-agglu
549	6	0.8	221	2	C81099	hypothetical prote	622	6	0.8	240	1	WMNVP6	p26 protein - Auto
550	6	0.8	221	2	C81842	hypothetical prote	623	6	0.8	240	2	T24363	hypothetical prote
551	6	0.8	221	2	G22322	glutaredoxin-relat	624	6	0.8	240	2	T41874	p26 protein (simil
552	6	0.8	221	2	C64040	hypothetical prote	625	6	0.8	241	2	S47673	leucine transport
553	6	0.8	221	2	F71668	hypothetical prote	626	6	0.8	241	2	E86012	ATP-binding compon
554	6	0.8	222	2	G95266	probable ABC trans	627	6	0.8	241	2	E91166	ATP-binding compon
555	6	0.8	222	2	A75063	flagellin B precur	628	6	0.8	241	2	S53812	EMGARA beta isofo
556	6	0.8	222	2	H89796	conserved hypotet	629	6	0.8	241	2	T25886	hypothetical prote
557	6	0.8	222	2	C97822	hypothetical prote	630	6	0.8	241	2	F91287	hypothetical prote
558	6	0.8	223	2	G83498	hypothetical prote	631	6	0.8	241	2	A86129	hypothetical prote
559	6	0.8	223	2	T35665	hypothetical prote	632	6	0.8	241	2	S56536	hypothetical prote
560	6	0.8	223	2	T29976	hypothetical prote	633	6	0.8	242	2	A45724	pectate lyase [EC
561	6	0.8	224	2	A83569	ribulose-phosphate	634	6	0.8	242	2	A86189	protein T25820.7 [
562	6	0.8	224	2	AF3646	1-fucose phosphate	635	6	0.8	242	2	B70570	hypothetical prote
563	6	0.8	224	2	S44964	ImbO protein - Str	636	6	0.8	243	2	F71861	hypothetical prote
564	6	0.8	224	2	AG1795	hypothetical prote	637	6	0.8	243	2	D64568	conserved hypotet
565	6	0.8	224	2	F84239	hypothetical prote	638	6	0.8	243	2	D90549	ribonuclease hii (
566	6	0.8	225	2	F76473	probable 1-acylgly	639	6	0.8	243	2	D83504	probable transcrip
567	6	0.8	225	2	A99206	hypothetical prote	640	6	0.8	243	2	S74031	hypothetical prote
568	6	0.8	225	2	AH3080	RntB family transp	641	6	0.8	243	2	AC3600	cellulase [EC 3.2.
569	6	0.8	225	2	AE2539	hypothetical prote	642	6	0.8	244	2	JC4708	gelatin-binding 28
570	6	0.8	226	2	C87637	conserved hypotet	643	6	0.8	244	2	T17916	hypothetical prote
571	6	0.8	226	2	H87291	conserved hypotet	644	6	0.8	245	2	JE0213	hisa protein - Cor
572	6	0.8	226	2	T43814	conserved hypotet	645	6	0.8	246	2	S47805	hypothetical 27.4K
573	6	0.8	227	1	MNIVX1	nonstructural prot	646	6	0.8	246	2	D91186	probable outer mem
574	6	0.8	227	1	MNIVX3	nonstructural prot	647	6	0.8	246	2	C86033	probable outer mem
575	6	0.8	227	1	MNIVX5	nonstructural prot	648	6	0.8	246	2	E70861	hypothetical prote
576	6	0.8	227	1	MNIVX9	nonstructural prot	649	6	0.8	246	2	C97177	C-terminal domain
577	6	0.8	227	1	MNIVX7	nonstructural prot	650	6	0.8	248	2	S34995	surface lipoprotei
578	6	0.8	227	1	I54426	MHC H2-w28-E alpha	651	6	0.8	249	2	T04436	ankyrin 3 homolog
579	6	0.8	227	2	E90432	triosephosphate is	652	6	0.8	250	2	B71859	probable oxidoredu
580	6	0.8	228	2	D82874	tRNA (Guanine-N1) -	653	6	0.8	250	2	E64564	short chain alcoho
581	6	0.8	228	2	T22219	hypothetical prote	654	6	0.8	250	2	C83678	hypothetical prote
582	6	0.8	229	2	B69409	conserved hypotet	655	6	0.8	250	2	AG3276	precorrin-3 methyl
583	6	0.8	229	2	T33141	hypothetical prote	656	6	0.8	251	2	AG3587	SN-glycerol-3-phos
584	6	0.8	230	1	VG1HBC	E1 membrane glycop	657	6	0.8	251	2	T03535	precorrin-6x reduc
585	6	0.8	230	1	QJ1749	E1 membrane glycop	658	6	0.8	251	2	G83188	undecaprenyl pyrop
586	6	0.8	230	1	MNIV1	nonstructural prot	659	6	0.8	251	2	S41587	hemd protein - Pse
587	6	0.8	230	1	MNIV1F	nonstructural prot	660	6	0.8	251	2	A82989	uroporphyrinogen-I
588	6	0.8	230	1	MNIV1A1	nonstructural prot	661	6	0.8	251	2	T35258	probable transcrip
589	6	0.8	230	1	MNIV1A2	nonstructural prot	662	6	0.8	252	1	A75073	phosphoesterase-re
590	6	0.8	230	1	MNIV1A3	nonstructural prot	663	6	0.8	252	2	S69786	phgK protein - Sal
591	6	0.8	230	1	MNIV1A4	nonstructural prot	664	6	0.8	252	2	AC0849	pathogenicity 1 is
592	6	0.8	230	2	QJ1172	membrane protein -	665	6	0.8	252	2	H82574	phage-related prot
593	6	0.8	230	2	S09648	nonstructural prot	666	6	0.8	253	2	F71667	hypothetical prote
594	6	0.8	230	2	D97697	exu regulon regula	667	6	0.8	253	2	D97820	zinc/manganese ABC
595	6	0.8	230	2	A75260	hypothetical prote	668	6	0.8	253	2	T30928	hypothetical prote
596	6	0.8	230	2	AD2923	transcription regu	669	6	0.8	254	2	H69057	hypothetical prote
597	6	0.8	231	2	H84478	hypothetical prote	670	6	0.8	254	2	E90969	hypothetical prote
598	6	0.8	232	2	T70699	probable paba prot	671	6	0.8	254	2	G69495	transcription regu
599	6	0.8	232	2	C10008	probable p-aminobe	672	6	0.8	254	2	D85742	hypothetical prote
600	6	0.8	232	2	D71218	hypothetical prote	673	6	0.8	255	1	HLMSD	H-2 class II histo
601	6	0.8	232	2	T15146	hypothetical prote	674	6	0.8	255	1	HLMSRA	H-2 class II histo
602	6	0.8	232	2	A61045	homeotic protein T	675	6	0.8	255	2	S65032	H4-transporing tw
603	6	0.8	232	2	B71325	probable V-type AT	676	6	0.8	255	2	D87355	RNA pseudouridylat
604	6	0.8	233	1	D69630	menaquinone biosyn	677	6	0.8	255	2	G64320	hypothetical prote
605	6	0.8	233	2	D75185	hypothetical prote	678	6	0.8	255	2	H86321	hypothetical prote
606	6	0.8	233	2	G83061	hypothetical prote	679	6	0.8	256	2	B55218	gramicidin S biosy
607	6	0.8	234	2	T20933	hypothetical prote	680	6	0.8	256	2	S58743	H4-transporing tw
608	6	0.8	234	2	H84297	phosphoglycerate d	681	6	0.8	256	2	AH2785	stationary-phase s
609	6	0.8	235	2	AG2218	hypothetical prote	682	6	0.8	256	2	B97565	stationary-phase s
610	6	0.8	236	2	E97723	hypothetical prote	683	6	0.8	256	2	G96774	hypothetical prote
611	6	0.8	236	2	A53853	apolipoprotein B m	684	6	0.8	256	2	T35290	probable transfera
612	6	0.8	236	2	D71548	hypothetical prote	685	6	0.8	256	2	F70812	probable lpqR prot
613	6	0.8	237	1	MNIV14	nonstructural prot	686	6	0.8	256	2	B82097	conserved hypotet

687 256 2 H82235 response regulator
688 1 S39747 ywfn protein - Bac
689 258 1 E84427 hypothetical prote
690 260 2 T18554 integral membrane
691 260 2 A87571 hypothetical prote
692 261 1 WNA087 early E1A 28K prot
693 261 2 E83199 conserved hypothet
694 262 2 T28643 Y4JP protein - Rhi
695 262 2 C95193 cell division prot
696 262 2 E87595 hydrolase, alpha/b
697 262 2 E83546 probable transcrip
698 264 2 A75076 membrane protein P
699 264 2 T47468 hypothetical prote
700 265 2 E83136 ferric enterobacti
701 265 2 D83577 conserved hypothet
702 265 2 B86338 protein F5M15.22 [I
703 265 2 T24623 hypothetical prote
704 266 1 U50344 tryptophan synthas
705 266 2 H82387 glucosamine-6-phos
706 266 2 T52322 chlorophyll a/b-bi
707 266 2 G83874 oligopeptide ABC t
708 266 2 H98059 hypothetical prote
709 266 2 AC2383 hypothetical prote
710 267 2 C89818 dihydroperate sy
711 267 2 B92220 1SU ribosomal prot
712 267 2 H87469 Th1J/Fip1 family p
713 267 2 A81280 probable sigma fac
714 268 1 A30584 interleukin-1 beta
715 268 2 T46749 hemolysin tlyC [va
716 268 2 D72564 probable myo-inosi
717 269 1 A46330 polyhedrin - Euxoa
718 270 2 S47479 outer surface prot
719 270 2 S32937 ureD protein - Kle
720 270 2 A42887 urease-associated
721 270 2 B69781 multidrug-efflux t
722 270 2 G71632 hypothetical prote
723 270 2 T19033 hypothetical prote
724 270 2 S62150 hypothetical prote
725 271 2 T32252 hypothetical prote
726 272 2 B72641 hypothetical prote
727 272 2 S71541 outer surface prot
728 273 2 S71544 outer surface prot
729 273 2 S71542 outer surface prot
730 273 2 I40102 outer surface prot
731 273 2 I40099 outer surface prot
732 273 2 I40097 outer surface prot
733 273 2 S23112 outer surface prot
734 273 2 S71531 outer surface prot
735 273 2 D90819 probable antitermi
736 273 2 B85627 probable antitermi
737 274 2 S83643 cytochrome-c oxida
738 274 2 B71964 enoyl-acyl carrier
739 275 2 C82717 50S ribosomal prot
740 275 2 T46825 catechol 1,2-dioxy
741 275 2 F82325 conserved hypothet
742 275 2 G87125 probable TetR-fam
743 276 2 B86163 protein F15K3.21 [I
744 276 2 T18742 hypothetical prote
745 276 2 A70425 conserved hypothet
746 276 2 AH0460 conserved hypothet
747 276 2 AG2807 hypothetical prote
748 276 2 E97586 probable noem gene
749 276 2 AG1842 uroporphyrinogen-I
750 276 2 B38965 hypothetical prote
751 277 2 A97530 BHI246 hypothetical
752 277 2 AB2749 ABC transporter, m
753 277 2 T49784 hypothetical prote
754 277 2 I60122 rsu-1 homolog - hu
755 277 2 S55770 RSP-1 protein - mo
756 278 2 T03942 rRNA N-glycosidase
757 278 2 T11550 probable membrane
758 278 2 F86679 hypothetical prote
759 278 2 A72411 hypothetical prote

760 0.8 6 278 2 E72307 transposase-relate
761 0.8 6 279 2 C82915 ribosomal protein
762 0.8 6 279 2 S76797 hypothetical prote
763 0.8 6 279 2 G83041 probable N-hydroxy
764 0.8 6 280 2 T03559 2-hydroxyhepta-2,4
765 0.8 6 280 2 D81705 conserved hypothet
766 0.8 6 280 2 T25898 conserved hypothet
767 0.8 6 280 2 H83523 conserved hypothet
768 0.8 6 281 1 R5YM2C ribosomal protein
769 0.8 6 282 2 S16617 opacity protein op
770 0.8 6 282 2 D64456 protein-export mem
771 0.8 6 282 2 S76911 hypothetical prote
772 0.8 6 283 2 C84321 hypothetical prote
773 0.8 6 283 2 D64448 hypothetical prote
774 0.8 6 283 2 AC1797 Partition protein
775 0.8 6 283 2 AE1423 Partition protein
776 0.8 6 284 2 G64207 ribosomal protein
777 0.8 6 284 2 A75356 conserved hypothet
778 0.8 6 284 2 E90190 conserved hypothet
779 0.8 6 284 2 S18957 fix23-5 protein -
780 0.8 6 284 2 AD3184 pirin-like protein
781 0.8 6 286 2 G83378 probable short-cha
782 0.8 6 288 2 A86384 unknown protein [i
783 0.8 6 288 2 I51620 homeobox protein -
784 0.8 6 288 2 C72407 hypothetical prote
785 0.8 6 289 2 E64330 dihydrodipicolinat
786 0.8 6 289 2 C97239 probable membrane
787 0.8 6 289 2 E90397 dihydrodipicolinat
788 0.8 6 290 2 T23416 hypothetical prote
789 0.8 6 290 2 H83467 probable transcrip
790 0.8 6 291 2 H81024 Arp synthase Fl, 9
791 0.8 6 291 2 H81970 Hs-transferrin tw
792 0.8 6 291 2 S66101 conserved hypothet
793 0.8 6 291 2 G83660 conserved hypothet
794 0.8 6 292 2 T06201 xyloglucan endo-1,
795 0.8 6 292 2 T26344 hypothetical prote
796 0.8 6 293 2 P95859 probable dihydrodi
797 0.8 6 293 2 A89818 hypothetical prote
798 0.8 6 293 2 D64984 hypothetical trans
799 0.8 6 293 2 A98010 probable transcrip
800 0.8 6 293 2 C85854 probable transcrip
801 0.8 6 294 2 B72246 dihydrodipicolinat
802 0.8 6 294 2 S72322 ribosomal protein
803 0.8 6 294 2 S05328 inner membrane pro
804 0.8 6 294 2 AE0780 cytidine deaminase
805 0.8 6 294 2 H96662 hypothetical prote
806 0.8 6 294 2 S19226 cold-regulated pro
807 0.8 6 295 2 C69180 adhesion protein -
808 0.8 6 295 2 T22833 hypothetical prote
809 0.8 6 296 2 AH3343 serine O-acetyltra
810 0.8 6 296 2 A83109 hypothetical prote
811 0.8 6 296 2 E69500 molybdopterin bios
812 0.8 6 296 2 T29892 hypothetical prote
813 0.8 6 296 2 S52254 copper resistance
814 0.8 6 296 2 H95404 hypothetical prote
815 0.8 6 296 2 H75557 hypothetical prote
816 0.8 6 296 2 AD2434 hypothetical prote
817 0.8 6 297 2 T13264 probable sugar upt
818 0.8 6 297 2 T36724 repressor protein
819 0.8 6 297 2 F69152 probable membrane
820 0.8 6 298 1 F69152 ribokinase - Metha
821 0.8 6 298 2 S53761 triose-phosphate i
822 0.8 6 298 2 D75481 aldose epimerase i
823 0.8 6 299 2 B95149 heat shock protein
824 0.8 6 299 2 F72210 hypothetical prote
825 0.8 6 299 2 T26365 hypothetical prote
826 0.8 6 299 2 AE3421 peptidoglycan bind
827 0.8 6 299 2 AE3153 transcription regu
828 0.8 6 300 1 S40827 probable sugar kin
829 0.8 6 300 1 H70557 probable acyl-CoA
830 0.8 6 300 2 T01598 hypothetical prote
831 0.8 6 300 2 S69028 probable membrane
832 0.8 6 301 1 RLZMFI rRNA N-glycosidase

833	6	0.8	302	2	A93017	heat shock protein	906	6	0.8	321	2	C84664	epoxide hydrolase
834	6	0.8	302	2	S11860	rRNA N-glycosidase	907	6	0.8	321	2	T07043	probable epoxide h
835	6	0.8	302	2	D82996	probable transcrip	908	6	0.8	321	2	JC5460	intracellular alka
836	6	0.8	302	2	T48871	catechol 1,2-dioxy	909	6	0.8	321	2	F82296	conserved hypotet
837	6	0.8	302	2	S50609	hypothetical prote	910	6	0.8	321	2	H83560	probable type II s
838	6	0.8	302	2	S34390	hypothetical prote	911	6	0.8	322	2	G83922	intracellular alka
839	6	0.8	302	2	A90572	aquaporin-like tra	912	6	0.8	322	2	G82018	lipopolysaccharide
840	6	0.8	302	2	D83268	conserved hypotet	913	6	0.8	322	2	B81001	heptosyltransferas
841	6	0.8	302	2	A95387	protein imported	914	6	0.8	322	2	T14597	proteinase homolog
842	6	0.8	303	1	A55592	cheA activity-modu	915	6	0.8	323	2	A55983	microtubule-associ
843	6	0.8	303	2	H97212	co/Zn/cd efflux sy	916	6	0.8	323	2	E36144	cobd protein - pse
844	6	0.8	304	2	S11859	rRNA N-glycosidase	917	6	0.8	323	2	S47966	probable lipid tra
845	6	0.8	304	2	T18005	hypothetical prote	918	6	0.8	323	2	D82987	hypothetical prote
846	6	0.8	304	2	E70698	hypothetical prote	919	6	0.8	324	2	A46631	lactose-binding le
847	6	0.8	305	2	S23711	oligofactory factor	920	6	0.8	324	2	T31992	hypothetical prote
848	6	0.8	305	2	D47099	nitrogen assimilat	921	6	0.8	324	2	T49143	hypothetical prote
849	6	0.8	305	2	H90976	nitrogen assimilat	922	6	0.8	324	2	T18763	hypothetical prote
850	6	0.8	305	2	F85823	nitrogen assimilat	923	6	0.8	324	2	A31920	collagen sgt-1 pre
851	6	0.8	305	2	D64963	nitrogen assimilat	924	6	0.8	324	2	S48122	transcription fact
852	6	0.8	306	2	T50038	beta-1,3-N-acetylgl	925	6	0.8	325	2	E83387	copper resistance
853	6	0.8	306	2	T70657	probable mr - Myc	926	6	0.8	325	2	A82281	ferric vibriobacti
854	6	0.8	307	2	AH3112	hypothetical prote	927	6	0.8	325	2	H70073	two-component sens
855	6	0.8	307	2	E98174	oligopeptide ABC t	928	6	0.8	325	2	T36562	probable ion-trans
856	6	0.8	308	2	D82897	conserved hypotet	929	6	0.8	326	2	T04344	peroxidase (EC 1.1
857	6	0.8	308	2	E64534	conserved hypotet	930	6	0.8	326	2	AC0878	Type II secretion,
858	6	0.8	308	2	C71972	hypothetical prote	931	6	0.8	327	2	S53305	ferredoxin-NADP re
859	6	0.8	309	2	S76393	hypothetical prote	932	6	0.8	327	2	AC0156	probable cobalamin
860	6	0.8	309	2	D70464	hypothetical prote	933	6	0.8	327	2	T18766	hypothetical prote
861	6	0.8	309	2	AH1528	conserved hypotet	934	6	0.8	327	2	T35482	hypothetical prote
862	6	0.8	309	2	AF1171	conserved hypotet	935	6	0.8	327	2	F72482	hypothetical prote
863	6	0.8	309	2	E75565	hypothetical prote	936	6	0.8	327	2	F95889	probable dehydroge
864	6	0.8	309	2	T36297	probable DNA-bind	937	6	0.8	328	2	AE3312	ribose-phosphate d
865	6	0.8	310	1	JU0299	ADPGlyceronamo-he	938	6	0.8	328	2	A91012	hypothetical prote
866	6	0.8	310	2	E90231	UDP-glucose 4-epim	939	6	0.8	328	2	C85856	hypothetical prote
867	6	0.8	310	2	A98191	ADP-L-glycero-D-ma	940	6	0.8	328	2	D64986	yeir protein - Esc
868	6	0.8	310	2	B86038	ADP-L-glycero-D-ma	941	6	0.8	328	2	AD0784	conserved hypotet
869	6	0.8	310	2	F90011	conserved hypotet	942	6	0.8	328	2	T19815	hypothetical prote
870	6	0.8	311	1	D69219	integrase-recombin	943	6	0.8	328	2	T42996	hypothetical prote
871	6	0.8	311	2	AF2668	deacetylase [impor	944	6	0.8	329	1	A36952	hypothetical prote
872	6	0.8	311	2	D97450	histone deacetylase	945	6	0.8	329	2	AE0378	CDP-6-deoxy-delta
873	6	0.8	311	2	F86341	hypothetical prote	946	6	0.8	329	2	F82922	phosphate transpor
874	6	0.8	312	1	G87328	ISCC2, transposase	947	6	0.8	329	2	C75400	geranylgeranyl dip
875	6	0.8	312	1	Z2BPB4	rIIB protein - pha	948	6	0.8	329	2	T44459	arginine metabolism
876	6	0.8	312	2	T35400	probable phytoene	949	6	0.8	329	2	A83405	probable hydroxyla
877	6	0.8	313	2	A64084	lysophospholipase	950	6	0.8	329	2	AD1083	dinitrogenase redu
878	6	0.8	313	2	D95866	probable transcrip	951	6	0.8	330	2	G70792	hypothetical prote
879	6	0.8	313	2	T21966	hypothetical prote	952	6	0.8	330	2	B70849	hypothetical prote
880	6	0.8	313	2	AF1973	hypothetical prote	953	6	0.8	330	2	S39588	peptide transport
881	6	0.8	314	2	T36913	probable integral	954	6	0.8	330	2	AH0656	peptide transport
882	6	0.8	314	2	D87576	oxidoreductase, al	955	6	0.8	330	2	S35439	transcription fact
883	6	0.8	314	2	D90457	cytochrome b558/56	956	6	0.8	330	2	AE2535	hypothetical prote
884	6	0.8	314	2	T28879	hypothetical prote	957	6	0.8	331	1	DECHLM	L-lactate dehydrog
885	6	0.8	315	2	AG1131	B. subtilis Yeac h	958	6	0.8	331	2	F88638	protein F58F6.6 [i
886	6	0.8	315	2	AG1491	conserved hypotet	959	6	0.8	331	2	G86671	hypothetical prote
887	6	0.8	315	2	JCS201	chemoreceptor TB56	960	6	0.8	331	2	T28208	hypothetical prote
888	6	0.8	315	2	AD3547	ribose transport s	961	6	0.8	332	2	S12151	L-lactate dehydrog
889	6	0.8	315	2	AF3060	conserved hypotet	962	6	0.8	332	2	B86176	protein F19P19.13
890	6	0.8	315	2	E98226	hypothetical prote	963	6	0.8	332	2	T12433	malate dehydrogena
891	6	0.8	315	2	B83874	nickel ABC transpo	964	6	0.8	332	2	T02935	malate dehydrogena
892	6	0.8	316	2	T34023	hypothetical prote	965	6	0.8	332	2	T09291	malate dehydrogena
893	6	0.8	317	2	H75453	electron transfer	966	6	0.8	332	2	T23903	hypothetical prote
894	6	0.8	317	2	T39869	probable lysophosp	967	6	0.8	332	2	B81383	probable lipoprote
895	6	0.8	318	2	T49714	related to spliceo	968	6	0.8	333	2	E95212	hypothetical prote
896	6	0.8	319	1	ERADF3	fiber protein - hu	969	6	0.8	333	2	E98076	catabolite control
897	6	0.8	319	2	I60446	Shiga-like cytotox	970	6	0.8	334	2	T36052	probable cytochrom
898	6	0.8	319	2	S52807	fiber protein - hu	971	6	0.8	334	2	H69148	hypothetical prote
899	6	0.8	319	2	A85437	hypothetical prote	972	6	0.8	334	2	E69361	signal-transducing
900	6	0.8	320	1	DAAL2E	catechol 2,3-dioxy	973	6	0.8	334	2	AG3029	hypothetical prote
901	6	0.8	320	2	D84664	probable epoxide h	974	6	0.8	334	2	B87545	hypothetical prote
902	6	0.8	320	2	B82520	hypothetical prote	975	6	0.8	334	2	E86794	hypothetical prote
903	6	0.8	320	2	AB3630	taurine-binding pe	976	6	0.8	334	2	G83533	hypothetical prote
904	6	0.8	321	1	S58614	cytochrome c-type	977	6	0.8	334	2	T20562	hypothetical prote
905	6	0.8	321	1	JQ0288	cytochrome c-type	978	6	0.8	334	2	B97789	hypothetical prote

979 6 0.8 335 1 XXFOGA
980 6 0.8 335 2 S31625
981 6 0.8 335 2 S81736
982 6 0.8 335 2 T25498
983 6 0.8 335 2 T34296
984 6 0.8 336 1 S75272
985 6 0.8 336 2 H96799
986 6 0.8 336 2 B90071
987 6 0.8 336 2 T44988
988 6 0.8 336 2 B72653
989 6 0.8 337 2 T23591
990 6 0.8 338 2 S12731
991 6 0.8 338 2 D84249
992 6 0.8 338 2 B64148
993 6 0.8 338 2 C64389
994 6 0.8 339 2 B36868
995 6 0.8 339 2 T16273
996 6 0.8 340 1 OKBYR1
997 6 0.8 340 2 S69194
998 6 0.8 340 2 E70714
999 6 0.8 340 2 T32646
1000 6 0.8 340 2 T05120

ALIGNMENTS

RESULT 1
D81883
Probable outer-membrane receptor protein NMA1161 [imported] - Neisseria meningitidis (str
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: D81883
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: D81883
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-764 <PAR>
A;Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84423.1; PID:g737985
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA1161

Query Match 23.1%; Score 175; DB 2; Length 764;
Best Local Similarity 100.0%; Pred. No. 3.1e-174; Indels 0; Gaps 0;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 473 PLPLGAHQRTARFALSGNNYFTPOHKLSTASHQERLPSTQELYAHGKHVATNTEVG 532
Db 479 PLPLGAHQRTARFALSGNNYFTPOHKLSTASHQERLPSTQELYAHGKHVATNTEVG 538
Qy 533 NKHLNKSNNIELALGVEGRWQNLALYRNFYIYAQTLDNGRGPKEIDDSMKL 592
Db 539 NKHLNKSNNIELALGVEGRWQNLALYRNFYIYAQTLDNGRGPKEIDDSMKL 598
Qy 593 VRYNQSGADFYGAEGEYIFKPTPRYRIGVSGDYVGRGLKNLPSLPGRDAYGNRP 647
Db 599 VRYNQSGADFYGAEGEYIFKPTPRYRIGVSGDYVGRGLKNLPSLPGRDAYGNRP 653

RESULT 2
A72479
hypothetical protein APE2473 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: A72479
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aerop
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: A72479
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-205 <KAW>
A;Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81489.1; PID:g5106178
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE2473
C;Superfamily: Aeropyrum pernix hypothetical protein APE2473
Query Match 1.2%; Score 9; DB 2; Length 205;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 318 HAHAHNGKP 326
Db 162 HAHAHNGKP 170

RESULT 3
T05478
Peroxidase (EC 1.11.1.7) prxrl - Arabidopsis thaliana
N;Alternate names: protein T805.170
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 26-Aug-1999
C;Accession: T05478
R;Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schuel
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15417
A;Accession: T05478
A;Molecule type: DNA
A;Residues: 1-323 <BEV>
A;Cross-references: EMBL:AL021890
A;Experimental source: cultivar Columbia; BAC clone T805
C;Genetics:
A;Gene: prxrl
A;Map position: 4
A;Introns: 69/3; 132/3; 188/1
A;Note: T805.170
C;Superfamily: peroxidase
C;Keywords: oxidoreductase

Query Match 1.2%; Score 9; DB 2; Length 323;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 408 ALSATSEAV 416
Db 9 ALSATSEAV 17

RESULT 4
S54307
myosin heavy chain - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 02-Mar-2001
C;Accession: S54307
R;Reinhard, J.; Scheel, A.A.; Diekmann, D.; Hall, A.; Ruppert, C.; Baehler, M.
EMBO J. 14, 697-704, 1995
A;Title: A novel type of myosin implicated in signalling by rho family GTPases.
A;Reference number: S54307; MUID:95188874; PMID:7882973
A;Accession: S54307
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1980 <REI>
A;Cross-references: EMBL:X77609; NID:g639998; PIDN:CAA54700.1; PID:g639999
C;Superfamily: rat myosin heavy chain; myosin motor domain homology; protein kinase C
F;149-942/Domain: myosin motor domain homology <MMOT>
F;239-246/Region: nucleotide-binding motif A (P-loop)
F;1593-1641/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 1.2%; Score 9; DB 2; Length 1980;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TEGLYRKSG 213
 |||||
 Db 1690 TEGLYRKSG 1698

RESULT 5
 A59256
 myosin-IXb [similarity] - human
 C:Species: Homo sapiens (man)
 C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 02-Mar-2001
 C:Accession: A59256; I61700
 R:Wirth, J.A.; Jensen, K.A.; Post, P.L.; Bement, W.M.; Mooseker, M.S.
 J. Cell Sci. 109, 653-661, 1996
 A:Title: Human myosin-IXb, an unconventional myosin with a chimerin-like rho/rac GTPase-
 A:Reference number: A59256; MUID:97063843; PMID:8907710
 A:Accession: A59256
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-2022 <WIR>
 A:Cross-references: GB:U42391; NID:g1147782; PIDN:AAC50402.1; PID:g1147783
 R:Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.
 Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
 A:Title: Identification and overlapping expression of multiple unconventional myosin gen
 A:Reference number: A55758; MUID:94294418; PMID:8022818
 A:Accession: I61700
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 234-322 <RES>
 A:Cross-references: GB:I29149; NID:g457257; PIDN:AAA20912.1; PID:g531142
 C:Genetics:
 A:Gene: GDB:MYO9B; OMIM:602129
 A:Map position: 19p13.1
 C:Superfamily: rat myosin heavy chain; myosin motor domain homology; protein kinase C zi
 C:Keywords: nucleotide binding; P-loop
 F:149-941/Domain: myosin motor domain homology #status atypical <WMO>
 F:239-246/Region: nucleotide-binding motif A (P-loop)

Query Match 1.2%; Score 9; DB 2; Length 2022;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 TEGLYRKSG 213
 |||||
 Db 1730 TEGLYRKSG 1738

RESULT 6
 AB1484
 hypothetical protein lin0409 [imported] - Listeria innocua (strain Clip11262)
 C:Species: Listeria innocua
 C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AB1484
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AB1484
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-99 <GUA>
 A:Cross-references: GB:AL592022; PIDN:CAC95642.1; PID:g16412838; GSPDB:GN00178
 A:Experimental source: strain Clip11262
 C:Genetics:
 A:Gene: lin0409

Query Match 1.1%; Score 8; DB 2; Length 99;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTTLKPI 9
 |||||
 Db 46 AQTTLKPI 53

RESULT 7
 AH1123
 hypothetical protein lmo0391 [imported] - Listeria monocytogenes (strain EGD-e)
 C:Species: Listeria monocytogenes
 C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AH1123
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AH1123
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-99 <GUA>
 A:Cross-references: GB:NC_003210; PIDN:CAC98470.1; PID:g16409769; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo0391

Query Match 1.1%; Score 8; DB 2; Length 99;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTTLKPI 9
 |||||
 Db 46 AQTTLKPI 53

RESULT 8
 AE2883
 nitroreductase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AE2883
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCle
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AE2883
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-196 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAL43483.1; PID:g17740989; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu2496
 A:Map position: circular chromosome

Query Match 1.1%; Score 8; DB 2; Length 196;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 LSSGNLEK 187
 |||||
 Db 70 LSSGNLEK 77

```

Bioorg. Khim. 16, 324-335, 1990
A;Title: NAD-dependent formate dehydrogenase from methylotrophic bacterium Pseudomonas
A;Reference number: JU0334; MUID:90290536; PMID:2357236
A;Accession: JU0334
A;Molecule type: protein
A;Residues: 1-393 <POP>
A;Note: article in Russian with English abstract
C;Superfamily: Neurospora formate dehydrogenase
C;Keywords: homodimer; NAD; oxidoreductase

Query Match 1.1%; Score 8; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 VSGELGLR 179
Db 53 VSGELGLR 60

RESULT 12
T19180
hypothetical protein C10C5.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19180
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19085
A;Accession: T19180
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-399 <WIL>
A;Cross-references: EMBL:Z68214; PIDN:CAA92445.1; GSPDB:GN00022; CESP:C10C5.3
A;Experimental source: clone C10C5
C;Genetics:
A;Gene: CESP:C10C5.3
A;Map position: 4
A;Introns: 27/1; 48/3; 215/3; 268/1; 303/3; 330/2; 350/3; 383/1

Query Match 1.1%; Score 8; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 LINTPLLA 22
Db 361 LINTPLLA 368

RESULT 13
JC7815
formate dehydrogenase (EC 1.2.1.2) - Paracoccus sp. (Strain 12-A)
C;Species: Paracoccus sp.
C;Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 03-Jun-2002
C;Accession: JC7815
R;Shinoda, T.; Satoh, T.; Mineki, S.; Iida, M.; Taguchi, H.
Biosci. Biotechnol. Biochem. 66, 271-276, 2002
A;Title: Cloning, nucleotide sequencing, and expression in Escherichia coli of the gen
A;Reference number: JC7815; PMID:1199398; MUID:21994041
A;Accession: JC7815
A;Molecule type: DNA
A;Residues: 1-400 <SHI>
A;Cross-references: DDBJ:AB071373
C;Comment: This enzyme, which catalyzes the conversion of formate to carbon dioxide wi
ti-enzyme systems, such as bioreactors involving NADH as a coenzyme, and plays a key r
C;Genetics:
A;Gene: fdh
C;Keywords: oxidoreductase

Query Match 1.1%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 VSGELGLR 179

```

```

probable enzyme (AE005300) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: C97659
R;Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: C97659
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88228.1; PID:G15157682; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4536
A;Map position: circular chromosome

Query Match 1.1%; Score 8; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 180 LSSGNLEK 187
Db 70 LSSGNLEK 77

RESULT 10
FNBSLC
beta-lactamase (EC 3.5.2.6) III precursor - Bacillus cereus
N;Alternate names: penicillinase III
C;Species: Bacillus cereus
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 12-Apr-1996
C;Accession: A27755
R;Hussain, M.; Pastor, F.I.J.; Lampen, J.O.
J. Bacteriol. 169, 579-586, 1987
A;Title: Cloning and sequencing of the blaZ gene encoding beta-lactamase III, a lipoprot
A;Reference number: A27755; MUID:87109042; PMID:3027036
A;Accession: A27755
A;Molecule type: DNA
A;Residues: 1-316 <HSN>
A;Experimental source: strain 569/H
C;Comment: This membrane-bound enzyme resembles several gram-positive class A beta-lacta
C;Comment: This enzyme is an acyl-glyceride thioether-linked lipoprotein with the hydrog
C;Genetics:
A;Gene: blaZ
C;Superfamily: beta-lactamase I
C;Keywords: antibiotic resistance; hydrolase; lipoprotein; membrane protein; penicillin
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-316/Product: beta-lactamase III #status predicted <MAT>
F;95/Active site: Ser #status predicted

Query Match 1.1%; Score 8; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 KPVLVLSIL 14
Db 278 KPVLVLSIL 285

RESULT 11
JU0334
formate dehydrogenase (EC 1.2.1.2) - Pseudomonas sp.
N;Alternate names: NAD-dependent formate dehydrogenase
C;Species: Pseudomonas sp.
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Sep-2002
C;Accession: JU0334
R;Popov, V.O.; Shumilin, I.A.; Ustinnikova, T.B.; Lamzin, V.S.; Egorov, T.A.

```



```

Db      54 VSGELGLR 61
|||||
RESULT 14
C95293
Probable NAD-dependent formate dehydrogenase [imported] - Sinorhizobium meliloti (strain
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 23-Sep-2002
C:Accession: C95293
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: C95293
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <KUR>
A:Cross-references: GB:AE006469; PIDN:AKG4909.1; PID:gl14523329; GSPDB:GNO0165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma0478
A:Genome: plasmid
C:Superfamily: Neurospora formate dehydrogenase

Query Match      1.1%; Score 8; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 13;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      172 VSGELGLR 179
|||||
Db      56 VSGELGLR 63

RESULT 15
S76859
hypothetical protein sll1550 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S76859
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76859
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-544 <KAN>
A:Cross-references: EMBL:D90917; GB:AB001339; NID:gl653836; PIDN:BA18771.1; PID:gl65386
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Start codon: GTG
C:Superfamily: Synechocystis hypothetical protein slr0042

Query Match      1.1%; Score 8; DB 2; Length 544;
Best Local Similarity 100.0%; Pred. No. 18;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      179 RLSSGNLE 186
|||||

Db      222 RLSSGNLE 229

RESULT 16
T40040
GTPase-activator protein for Rho-like GTPases - fission yeast (Schizosaccharomyces pom)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40040
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21901
A:Accession: T40040
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-892 <WOO>
A:Cross-references: EMBL:AL031517; PIDN:CAA20650.1; GSPDB:GNO0067; SPDB:SPBC38E12.03
A:Experimental source: strain 972h-; cosmid c28E12
C:Genetics:
A:Gene: SPDB:SPBC28E12.03
A:Map position: 2
A:Introns: 7/2; 124/1

Query Match      1.1%; Score 8; DB 2; Length 892;
Best Local Similarity 100.0%; Pred. No. 28;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      206 EGYLRKSG 213
|||||
Db      738 EGYLRKSG 745

RESULT 17
S61174
hypothetical protein YDR379w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D9481.4
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
C:Accession: S61174
R:Ding, H.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of S. cerevisiae cosmid 9481.
A:Reference number: S61159
A:Accession: S61174
A:Molecule type: DNA
A:Residues: 1-1009 <DIN>
A:Cross-references: EMBL:U28373; NID:g849184; PIDN:AAB64815.1; PID:g849200; MIPS:YDR37;
A:Experimental source: strain S288C (AB972)
C:Genetics:
A:Gene: SGD:RGA2
A:Cross-references: SGD:S0002787; MIPS:YDR379w
A:Map position: 4R
C:Superfamily: LIM metal-binding repeat homology
F:13-66/Domain: LIM metal-binding repeat homology <LIM2>

Query Match      1.1%; Score 8; DB 2; Length 1009;
Best Local Similarity 100.0%; Pred. No. 31;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      206 EGYLRKSG 213
|||||
Db      822 EGYLRKSG 829

RESULT 18
S34567
gene L protein - hemorrhagic septicemia virus (fragment)
C:Species: hemorrhagic septicemia virus
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995
C:Accession: S34567
R:Schuetze, H.
submitted to the EMBL Data Library, June 1993
A:Reference number: S34562

```

A;Accession: S34567
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-54 <SCH>
 A;Cross-references: EMBL:X73873
 C;Genetics:
 A;Gene: 1

Query Match 0.9%; Score 7; DB 2; Length 54;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 HQERLPS 513
 Db 11 HQERLPS 17
 |||||

RESULT 19
 A81783
 hypothetical protein NMA2114 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
 C;Species: Neisseria meningitidis
 C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 ; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A;Reference number: A81775; MUID:20222556; PMID:10761919
 A;Accession: A81783
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-76 <PAR>
 A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85328.1; PID:g738073
 A;Experimental source: serogroup A, strain Z2491
 C;Genetics:
 A;Gene: NMA2114

Query Match 0.9%; Score 7; DB 2; Length 76;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 DRIDANL 680
 Db 17 DRIDANL 23
 |||||

RESULT 20
 E69460
 conserved hypothetical protein AF1686 - Archaeoglobus fulgidus
 C;Species: Archaeoglobus fulgidus
 C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
 R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A;Reference number: A69250; MUID:98049343; PMID:9389475
 A;Accession: E69460
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-100 <KLE>
 A;Cross-references: GB:AE000987; GB:AE000782; NID:g2689310; PIDN:AAB89563.1; PID:g264886
 C;Superfamily: conserved hypothetical protein MJ0128

Query Match 0.9%; Score 7; DB 2; Length 100;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 NVAGLVD 157
 |||||

Db 57 NVAGLVD 63
 |||||

RESULT 21
 A72224
 hypothetical protein TM1659 - Thermotoga maritima (strain MSB8)
 C;Species: Thermotoga maritima
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C;Accession: A72224
 R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hick
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D
 C.M.
 Nature 399, 323-329, 1999
 A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome s
 A;Reference number: A72200; MUID:99287316; PMID:10360571
 A;Accession: A72224
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-108 <ARN>
 A;Cross-references: GB:AE001808; GB:AE000512; NID:g4982233; PIDN:AAD36726.1; PID:g4982
 A;Experimental source: strain MSB8
 C;Genetics:
 A;Gene: TM1659
 C;Superfamily: Thermotoga maritima hypothetical protein TM1659

Query Match 0.9%; Score 7; DB 2; Length 108;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 300 VDYDNGP 306
 Db 50 VDYDNGP 56
 |||||

RESULT 22
 S01185
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 3 - fruit fly (Drosophila melanoga
 C;Species: mitochondrion Drosophila melanogaster
 C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 03-Jun-2002
 C;Accession: S01185
 R;Garesse, R.
 Genetics 118, 649-663, 1988
 A;Title: Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary
 A;Reference number: S01185; MUID:88212147; PMID:3130291
 A;Accession: S01185
 A;Molecule type: DNA
 A;Residues: 1-117 <GAR>
 A;Cross-references: GB:M37275; EMBL:Y00610; NID:g336819; PIDN:AAA69709.1; PID:g894077
 A;Note: the author translated the initiation codon ATT for residue 1 as Ile
 C;Genetics:
 A;Gene: ND-3
 A;Cross-references: FlyBase:FBgn0013681
 A;Genome: mitochondrion
 A;Genetic code: SGC4
 A;Start codon: ATT
 C;Superfamily: NADH dehydrogenase (ubiquinone) chain 3
 C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 0.9%; Score 7; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 KALIDRE 467
 Db 28 KALIDRE 34
 |||||

RESULT 23
 G25797
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 3 - fruit fly (Drosophila yakuba)
 N;Alternate names: NADH-ubiquinone oxidoreductase chain 3
 C;Species: mitochondrion Drosophila yakuba
 C;Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 03-Jun-2002

C;Accession: G25797
R;Clary, D.O.; Wolstenholme, D.R.
J. Mol. Evol. 22, 252-271, 1985
A;Title: The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene
A;Reference number: A92962; MUID:86089137; PMID:3001325
A;Accession: G25797
A;Molecule type: DNA
A;Residues: 1-117 <CLA>
A;Cross-references: GB:X03240; GB:J01400; GB:J01402; GB:J01403; GB:J01406; GB:J01408; GB:
C;Genetics:
A;Gene: FlyBase:Dyak/mt:ND3
A;Cross-references: FlyBase:FBgn0013204
A;Genome: mitochondrion
A;Genetic code: GCG
A;Start codon: ATT
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 3
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 0.9%; Score 7; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 47; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 461 KALIDRE 467
|||||
DB 28 KALIDRE 34

RESULT 24
C87588
hypothetical protein CC2738 [imported] - *Caulobacter crescentus*
C;Species: *Caulobacter crescentus*
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
A;Accession: C87588
R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of *Caulobacter crescentus*.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87588
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-117 <STO>
A;Cross-references: GB:AE005673; NID:gl3424331; PIDN:AAK24703.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2738

Query Match 0.9%; Score 7; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 47; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 408 ALSATSE 414
|||||
DB 31 ALSATSE 37

RESULT 25
T05445
hypothetical protein F7K2.90 - *Arabidopsis thaliana*
C;Species: *Arabidopsis thaliana* (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
A;Accession: T05445
R;Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Meves, H.W.; Mayer, K.F.X.; Schueller
submitted to the Protein Sequence Database, November 1998
A;Reference number: Z15416
A;Cross-references: EMBL:AL033545
A;Molecule type: DNA
A;Residues: 1-125 <BEV>
A;Cross-references: EMBL:AL033545
A;Experimental source: cultivar Columbia; BAC clone F7K2
C;Genetics:
A;Map position: 4
A;Introns: 73/3

A;Note: F7K2.90

Query Match 0.9%; Score 7; DB 2; Length 125;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 501 LSLTASH 507
|||||
DB 8 LSLTASH 14

RESULT 26
T27080
hypothetical protein Y51A2D.10 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T27080
R;McMurray, A.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z20307
A;Accession: T27080
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-141 <WIL>
A;Cross-references: EMBL:AL021497; PIDN:CAA16408.1; GSPDB:GN00023; CESP:Y51A2D.10
A;Experimental source: clone Y51A2D
C;Genetics:
A;Gene: CESP:Y51A2D.10
A;Map position: 5
A;Introns: 54/1; 96/3
C;Superfamily: *Caenorhabditis* hypothetical protein C40H1.5

Query Match 0.9%; Score 7; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 56; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 35 TTVVVGK 41
|||||
DB 27 TTVVVGK 33

RESULT 27
T27071
hypothetical protein Y51A2D.9 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T27071
R;McMurray, A.
submitted to the EMBL Data Library, January 1998
A;Reference number: Z20307
A;Accession: T27071
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-141 <WIL>
A;Cross-references: EMBL:AL021497; PIDN:CAA16399.1; GSPDB:GN00023; CESP:Y51A2D.9
A;Experimental source: clone Y51A2D
C;Genetics:
A;Gene: CESP:Y51A2D.9
A;Map position: 5
A;Introns: 54/1; 96/3
C;Superfamily: *Caenorhabditis* hypothetical protein C40H1.5

Query Match 0.9%; Score 7; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 56; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 35 TTVVVGK 41
|||||
DB 27 TTVVVGK 33

RESULT 28
T35846

probable integral membrane protein - Streptomyces coelicolor (fragment)
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 03-Nov-2000
 C/Accession: T35846
 R/Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M. submitted to the EMBL Data Library, April 1999
 A/Reference number: Z21591
 A/Accession: T35846
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-149 <SAU>
 A/Cross-references: EMBL:AL049727; PIDN:CAB41547.1; GSPDB:GN000070; SCODEB:SC9B1.01c
 A/Experimental source: strain A3(2)
 C/Genetics:
 A/Gene: SCODEB:SC9B1.01c
 C/Superfamily: integral membrane protein HP0228

Query Match 0.9%; Score 7; DB 2; Length 149;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 258 DOYGLPA 264
 |||||
 Db 83 DOYGLPA 89

RESULT 29
 DTECR
 aspartate carbamoyltransferase (EC 2.1.3.2) regulatory chain - Escherichia coli (strain
 A/Alternate names: aspartate transcarbamylase; carbamylaspartotranskinase
 C/Species: Escherichia coli
 C/Date: 24-Apr-1984 #sequence_revision 03-Aug-1984 #text_change 01-Mar-2002
 C/Accession: A93985; A00560; S56470; I56404; G65236
 R/Schachman, H.K.; Pauza, C.D.; Navre, M.; Karels, M.J.; Wu, L.; Yang, Y.R. Proc. Natl. Acad. Sci. U.S.A. 81, 1115-1119, 1984
 A/Title: Location of amino acid alterations in mutants of aspartate transcarbamoylase: E
 A/Reference number: A93985; MUID:84119419; PMID:6364131
 A/Accession: A93985
 A/Molecule type: DNA
 A/Residues: 1-153 <SCH>
 A/Cross-references: GB:K01472; NID:gl47463; PIDN:AAA24477.1; PID:gl47465
 R/Weber, K. Nature 218, 1116-1119, 1968
 A/Title: New structural model of Escherichia coli aspartate transcarbamylase and the ami
 A/Reference number: A93154; MUID:68284659; PMID:4872216
 A/Accession: A00560
 A/Molecule type: protein
 A/Residues: 1-3, 'ND', 6-9, 'AE', 12-18, 'N', 20-23, 'E', 25-38, 'QD', 41-86, 'ND', 89-102, 'NID', 106
 R/Monaco, H.L.; Crawford, J.L.; Lipscomb, W.N. Proc. Natl. Acad. Sci. U.S.A. 75, 5276-5280, 1978
 A/Title: Three-dimensional structures of aspartate carbamoyltransferase from Escherichia
 A/Reference number: A93823; MUID:79074799; PMID:364472
 R/Xe, H.; Honzatko, R.B.; Lipscomb, W.N. Proc. Natl. Acad. Sci. U.S.A. 81, 4037-4040, 1984
 A/Title: Structure of unligated aspartate carbamoyltransferase of Escherichia coli at 2.
 A/Reference number: A93993; MUID:84248054; PMID:6377306
 A/Contents: annotation; X-ray crystallography, 2.6 angstroms; quaternary structure
 R/Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R. Nucleic Acids Res. 23, 2105-2119, 1995
 A/Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
 A/Reference number: S56314; MUID:95334362; PMID:7610040
 A/Accession: S56470
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-153 <BUR>
 A/Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97141.1; PID:g537086
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
 R/Cunin, R.; Jacobs, A.; Charlier, D.; Crabeel, M.; Herve, G.; Glansdorff, N.; Pierard, J. Mol. Biol. 186, 707-713, 1985
 A/Title: Structure-function relationship in allosteric aspartate carbamoyltransferase fr
 A/Reference number: I56404; MUID:86143826; PMID:3912513
 A/Accession: I56404

A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 128-153 <RES>
 A/Cross-references: GB:M28578; NID:g147480; PIDN:AAA24487.1; PID:g147481
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: G65236
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-153 <BLAT>
 A/Cross-references: GB:AE000495; GB:U00096; NID:g2367361; PIDN:AACT7201.1; PID:g179069
 A/Experimental source: strain K-12, substrain MG1655
 C/Genetics:
 A/Gene: pyrI
 A/Map position: 97 min
 A/Complex: heterododecamer of two trimers of catalytic chains and three dimers of regu
 C/Function:
 A/Description: catalyzes the transcarbamylation of carbanoyl phosphate and aspartate, t
 A/Pathway: pyrimidine nucleotide biosynthesis
 C/Superfamily: aspartate carbamoyltransferase regulatory chain
 C/Keywords: acyltransferase; heterododecamer; homodimer; homohehexamer; homotrimer; pyrI
 F109,114,138,141/Binding site: zinc (Cys) #status experimental

Query Match 0.9%; Score 7; DB 1; Length 153;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
 |||||
 Db 91 VVGKSRP 97

RESULT 30
 DTECR
 aspartate carbamoyltransferase (EC 2.1.3.2) regulatory chain - Salmonella typhimurium
 C/Species: Salmonella typhimurium
 C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 11-Jun-1999
 C/Accession: S00050
 R/Michaelis, G.; Kelln, R.A.; Nargang, F.E. Eur. J. Biochem. 166, 55-61, 1987
 A/Title: Cloning, nucleotide sequence and expression of the pyrBI operon of Salmonella
 A/Reference number: S00028; MUID:87246692; PMID:3036524
 A/Accession: S00050
 A/Molecule type: DNA
 A/Residues: 1-153 <MIC>
 A/Cross-references: GB:X05641; NID:g47861; PIDN:CAA29130.1; PID:g47864
 C/Comment: the active enzyme contains two trimers of catalytic chains and three dimers
 C/Genetics:
 A/Gene: pyrI
 C/Superfamily: aspartate carbamoyltransferase regulatory chain
 C/Keywords: heterododecamer; homodimer; homohehexamer; pyrimidine nucleotide biosynthesi
 F109,114,138,141/Binding site: zinc (Cys) #status predicted

Query Match 0.9%; Score 7; DB 1; Length 153;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
 |||||
 Db 91 VVGKSRP 97

RESULT 31
 E91281
 aspartate carbamoyltransferase (EC 2.1.3.2) regulatory chain - Escherichia coli (strai
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 31-Dec-2001
 C/Accession: E91281
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Iehi, K.; Yokoyama, K.; Han, C. Gasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic islands
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E91281
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA038644.1; PID:g13364698; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5221
C;Superfamily: aspartate carbamoyltransferase regulatory chain
C;Keywords: transferase

Query Match 0.9%; Score 7; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
Db 91 VVGKSRP 97
|||||

RESULT 32
F86122
aspartate carbamoyltransferase (EC 2.1.3.2) regulatory chain - *Escherichia coli* (strain
C;Species: *Escherichia coli*
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Dec-2001
C;Accession: F86122
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamoukis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206851
A;Accession: F86122
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <SNT>
A;Cross-references: GB:AE005174; NID:g12519252; PIDN:BA059442.1; GSPDB:GN00145; UWGP:Z58
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: pyrI
C;Superfamily: aspartate carbamoyltransferase regulatory chain
C;Keywords: transferase

Query Match 0.9%; Score 7; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
Db 91 VVGKSRP 97
|||||

RESULT 33
AC1059
aspartate carbamoyltransferase regulatory chain pyrI [imported] - *Salmonella enterica* subsp.
C;Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A;Note: this species has also been called *Salmonella typhi*
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC1059
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC1059
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-153 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD06920.1; PID:g16505568; GSPDB:GN00176

C;Genetics:
A;Gene: pyrI
C;Superfamily: aspartate carbamoyltransferase regulatory chain

Query Match 0.9%; Score 7; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 VVGKSRP 44
Db 91 VVGKSRP 97
|||||

RESULT 34
S76912
hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
C;Species: *Synechocystis* sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S76912
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocyst*
S.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76912
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-155 <KAN>
A;Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BA018824.1; PID:d1019
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

Query Match 0.9%; Score 7; DB 2; Length 155;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 SGNLEKL 188
Db 146 SGNLEKL 152
|||||

RESULT 35
T07089
dehydrin - soybean
C;Species: Glycine max (soybean)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Aug-1999
C;Accession: T07089
R;Hsu, T.F.; Tsai, F.Y.; Hsing, Y.I.; Chow, T.Y.
submitted to the EMBL Data Library, May 1997
A;Description: Glycine max mRNA for dehydrin.
A;Reference number: Z15912
A;Accession: T07089
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-166 <HSU>
A;Cross-references: EMBL:AF004807; NID:g2270989; PIDN:AA071225.1; PID:g2270990
A;Experimental source: strain Shi-shi; cotyledon
C;Genetics:
A;Gene: PM12
C;Superfamily: dehydrin-like protein

Query Match 0.9%; Score 7; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 88 YGGGASA 94
Db 59 YGGGASA 65
|||||

```

RESULT 36
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: C70228
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-173 <KLE>
A:Cross-references: GB:AE000794; NID:g2689981; PIDN:AA06368.1; PID:g2689985; TIGR:BBF
A:Experimental source: strain B31
C:Genetics:
A:Gene: plasmid

Query Match          0.9%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PIVLSIL 14
Db 154 PIVLSIL 160

RESULT 39
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87609
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-185 <STO>
A:Cross-references: GB:AE005673; NID:gi3424527; PIDN:AAK24869.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2907

Query Match          0.9%; Score 7; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 GDALGV 80
Db 122 GDALGV 128

RESULT 40
A:Title: The delta'-subunit of higher plant six-subunit mitochondrial F-1-ATPase is ho
A:Reference number: A41740; MUID:92112727; PMID:1370454
A:Accession: A41740
A:Molecule type: mRNA
A:Residues: 1-200 <MOR>
A:Cross-references: GB:D10660; GB:D90473; NID:g217937; PIDN:BA01511.1; PID:g217938
R:Kimura, T.; Nakamura, K.; Kajjura, H.; Hattori, H.; Nelson, N.; Asahi, T.
J. Biol. Chem. 264, 3183-3186, 1989
A:Title: Correspondence of minor subunits of plant mitochondrial F-1-ATPase to F-1F-0AT
A:Reference number: A33306; MUID:89123436; PMID:2536736
A:Accession: C33306
A:Molecule type: protein
A:Residues: 'X',23-38,'X',40-50,'X',52,'X',54,'X',56 <KIM>
A:Superfamily: H+-transporting ATP synthase epsilon chain
C:Keywords: hydrolase
F:1-21/Domain: transit peptide (mitochondrion) #status predicted <TRP>

```

```

RESULT 36
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: C70228
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-173 <KLE>
A:Cross-references: GB:AE000794; NID:g2689981; PIDN:AA06368.1; PID:g2689985; TIGR:BBF
A:Experimental source: strain B31
C:Genetics:
A:Gene: plasmid

Query Match          0.9%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PIVLSIL 14
Db 154 PIVLSIL 160

RESULT 39
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87609
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-185 <STO>
A:Cross-references: GB:AE005673; NID:gi3424527; PIDN:AAK24869.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2907

Query Match          0.9%; Score 7; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 GDALGV 80
Db 122 GDALGV 128

RESULT 40
A:Title: The delta'-subunit of higher plant six-subunit mitochondrial F-1-ATPase is ho
A:Reference number: A41740; MUID:92112727; PMID:1370454
A:Accession: A41740
A:Molecule type: mRNA
A:Residues: 1-200 <MOR>
A:Cross-references: GB:D10660; GB:D90473; NID:g217937; PIDN:BA01511.1; PID:g217938
R:Kimura, T.; Nakamura, K.; Kajjura, H.; Hattori, H.; Nelson, N.; Asahi, T.
J. Biol. Chem. 264, 3183-3186, 1989
A:Title: Correspondence of minor subunits of plant mitochondrial F-1-ATPase to F-1F-0AT
A:Reference number: A33306; MUID:89123436; PMID:2536736
A:Accession: C33306
A:Molecule type: protein
A:Residues: 'X',23-38,'X',40-50,'X',52,'X',54,'X',56 <KIM>
A:Superfamily: H+-transporting ATP synthase epsilon chain
C:Keywords: hydrolase
F:1-21/Domain: transit peptide (mitochondrion) #status predicted <TRP>

```

F:22-200/Product: H+-transporting ATP synthase delta' chain #status predicted <MAT>

Query Match 0.9%; Score 7; DB 2; Length 200;
Best Local Similarity 100.0%; Pred.No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 674 DRIDANL 680
Db 154 DRIDANL 160

Search completed: November 14, 2003, 11:08:47
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:04:59 ; Search time 42 Seconds

(without alignments)
4657.235 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 758

Sequence: 1 MAQTILKPIVLSILLINTPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table:

OLIGO Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rviro:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	286	37.7	758	16 Q9JZN9	Q9jzn9 neisseria m
2	175	23.1	764	16 Q9JUS2	Q9jus2 neisseria m
3	12	1.6	809	16 Q9CLW8	Q9clw8 pasteurella
4	9	1.2	205	17 Q9Y911	Q9y911 aeropyrum p
5	9	1.2	766	2 Q9AER9	Q9aer9 pasteurella
6	8	1.1	93	2 Q9ZH42	Q9zh42 morganella
7	8	1.1	99	16 Q9ZEP9	Q9zep9 listeria in
8	8	1.1	99	16 Q9YXK6	Q9yxk6 listeria mo
9	8	1.1	171	11 Q9DAA9	Q9daa9 mus musculus
10	8	1.1	268	5 Q9VNN6	Q9vnn6 drosophila
11	8	1.1	268	5 Q9NH68	Q9nh68 drosophila
12	8	1.1	398	2 Q9F7P9	Q9f7p9 uncultured
13	8	1.1	399	2 Q93UW1	Q93uw1 hyphomicrob
14	8	1.1	399	5 Q17898	Q17898 caenorhabdi
15	8	1.1	400	2 Q9R5V0	Q9r5v0 mycobacteri
16	8	1.1	400	2 Q93GW3	Q93gw3 paracoccus

17	8	1.1	401	2 Q93GV1	Q93gv1 mycobacteri
18	8	1.1	401	16 Q93OE7	Q93oe7 rhizobium m
19	8	1.1	402	2 Q08375	Q08375 moraxella s
20	8	1.1	427	2 Q9F771	Q9f771 pseudomonas
21	8	1.1	544	16 P74655	P74655 synecocyst
22	8	1.1	925	16 Q9CWR7	Q9cmr7 pasteurella
23	8	1.1	933	3 Q74360	Q74360 schizosacch
24	8	1.1	984	16 Q8PD15	Q8pd15 xanthomonas
25	8	1.1	1009	3 Q06407	Q06407 saccharomyc
26	8	1.1	2319	10 Q8H614	Q8h614 zea mays (m
27	8	1.1	5017	2 Q8GBX6	Q8gbx6 polyangium
28	7	0.9	48	2 Q8KY74	Q8ky74 mycobacteri
29	7	0.9	48	2 Q8KY83	Q8ky83 mycobacteri
30	7	0.9	48	2 Q8KHQ1	Q8khl1 mycobacteri
31	7	0.9	55	12 Q08482	Q08482 viral hemor
32	7	0.9	60	10 Q8L411	Q8l411 oryza sativ
33	7	0.9	76	16 Q9JSV8	Q9jsv8 neisseria m
34	7	0.9	86	2 Q8KY84	Q8ky84 mycobacteri
35	7	0.9	86	2 Q8KY71	Q8ky71 mycobacteri
36	7	0.9	86	2 Q8KY78	Q8ky78 mycobacteri
37	7	0.9	86	2 Q8KY80	Q8ky80 mycobacteri
38	7	0.9	86	2 Q8KY79	Q8ky79 mycobacteri
39	7	0.9	86	2 Q8KY77	Q8ky77 mycobacteri
40	7	0.9	91	5 Q9BK93	Q9bk93 cryptospori
41	7	0.9	91	5 Q9BK94	Q9bk94 cryptospori
42	7	0.9	94	16 Q8PMD7	Q8pmd7 xanthomonas
43	7	0.9	100	17 Q28587	Q28587 archaeoglob
44	7	0.9	105	2 Q9RZW7	Q9rzw7 borrelia bu
45	7	0.9	108	16 Q9X1Y9	Q9x1y9 thermotoga
46	7	0.9	117	8 Q9MD67	Q9md67 drosophila
47	7	0.9	117	8 Q9MGW8	Q9mgw8 drosophila
48	7	0.9	117	8 Q9MGL9	Q9mgl9 drosophila
49	7	0.9	117	16 Q9A4T9	Q9a4t9 caulobacter
50	7	0.9	124	12 Q9J521	Q9j521 fowlpox vir
51	7	0.9	125	10 Q9SUW8	Q9suw8 arabidopsis
52	7	0.9	138	13 Q9DBZ7	Q9dbz7 scyllorhinu
53	7	0.9	141	5 Q9XXR4	Q9xxr4 caenorhabdi
54	7	0.9	141	5 Q9XXQ6	Q9xxq6 caenorhabdi
55	7	0.9	141	10 Q8H5V0	Q8h5v0 oryza sativ
56	7	0.9	144	5 Q61831	Q61831 caenorhabdi
57	7	0.9	149	11 Q8C542	Q8c542 mus musculu
58	7	0.9	155	16 P74705	P74705 synecocyst
59	7	0.9	159	16 Q9KY20	Q9ky20 streptomyce
60	7	0.9	162	6 Q8SP19	Q8spt9 macaca mula
61	7	0.9	163	5 Q8MXV4	Q8mxv4 cryptospori
62	7	0.9	166	10 Q23957	Q23957 glycine max
63	7	0.9	169	5 Q9NC83	Q9nc83 strongyloce
64	7	0.9	171	16 Q8UCY5	Q8ucy5 agrobacteri
65	7	0.9	173	16 Q51013	Q51013 borrelia bu
66	7	0.9	184	16 Q8DD61	Q8dd61 vibrio vuln
67	7	0.9	185	16 Q9A4C9	Q9a4c9 caulobacter
68	7	0.9	189	16 Q98G88	Q98g88 rhizobium l
69	7	0.9	198	11 Q8CAH4	Q8cah4 mus musculu
70	7	0.9	201	16 Q9XQ03	Q9xq03 thermotoga
71	7	0.9	202	5 Q9U8Y6	Q9u8y6 tachypleus
72	7	0.9	202	5 Q9U8Y8	Q9u8y8 tachypleus
73	7	0.9	202	5 Q9U8Y7	Q9u8y7 tachypleus
74	7	0.9	202	17 Q9VEY6	Q9vey6 aeropyrum p
75	7	0.9	207	15 Q71842	Q71842 human immun
76	7	0.9	207	15 Q71841	Q71841 human immun
77	7	0.9	215	5 Q9TZD2	Q9tzd2 caenorhabdi
78	7	0.9	222	5 Q16539	Q16539 caenorhabdi
79	7	0.9	232	2 Q9R655	Q9r655 synecococc
80	7	0.9	232	2 Q8GAA2	Q8ga2 synecococc
81	7	0.9	232	4 Q8IX63	Q8ix63 homo sapien
82	7	0.9	234	16 Q8XIF6	Q8xif6 clostridium
83	7	0.9	234	17 Q97WB6	Q97wb6 sulfolobus
84	7	0.9	234	17 Q96XJ2	Q96xj2 sulfolobus
85	7	0.9	237	10 Q8RZY8	Q8rzy8 oryza sativ
86	7	0.9	241	16 Q99U19	Q99u19 staphylococ
87	7	0.9	241	16 Q8CSH9	Q8csh9 staphylococ
88	7	0.9	243	4 Q96CG8	Q96cgb homo sapien
89	7	0.9	243	16 Q8XMM3	Q8xmm3 clostridium

QY 593 VRYNSGADFYGAEGEIIYFKPTERYIGVSGDYVRGLKNLPSLPGRDAYGNRP 647
 |||||
 Db 599 VRYNSGADFYGAEGEIIYFKPTERYIGVSGDYVRGLKNLPSLPGRDAYGNRP 653

RESULT 3

Q9CLW8 PRELIMINARY; PRT; 809 AA.
 AC Q9CLW8; 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein PM1081.
 GN PM1081.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70; PubMed=11248100;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
 DR EMBL: AE006149; AAK03165.1; -
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_dep_Rcg; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 809 AA; 90903 MW; 23026B7DFCBF14DB CRC64;

Query Match 1.6%; Score 12; DB 16; Length 809;
 Best Local Similarity 100.0%; Pred. No. 0.0043; Mismatches 0; Indels 0; Gaps 0;
 Matches 12; Conservative 0;

QY 89 GGGASAPVIRGQ 100

Db 103 GGGASAPVIRGQ 114
 |||||

RESULT 4

Q9Y911 PRELIMINARY; PRT; 205 AA.
 AC Q9Y911; 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein APE2473.
 GN APE2473.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococcaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kavarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyma A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101 (1999).
 DR EMBL: AP000064; BAA81489.1; -
 DR InterPro: IPR004843; M-peptidase.
 DR Pfam: PF00149; Metallophos; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 205 AA; 23063 MW; 7A1D29E93615304F CRC64;

Query Match 1.2%; Score 9; DB 17; Length 205;

Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 318 HAHANGKP 326
 |||||
 Db 162 HAHANGKP 170

RESULT 5

Q9AER9 PRELIMINARY; PRT; 766 AA.
 AC Q9AER9; 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Iron-regulated outer membrane protein.
 GN IRP.
 OS Pasteurella haemolytica.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Mannheimia.
 OX NCBI_TaxID=75985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lo R.Y., Graham M.R.;
 RT "Putative TonB dependent receptor of Mannheimia haemolytica.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL: AY028475; AAK39743.1; -
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_dep_Rcg; 1.
 DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
 DR KW Membrane; Outer membrane; Receptor; TonB box.
 SQ SEQUENCE 766 AA; 86440 MW; 7D76FAE05D7C5619 CRC64;

Query Match 1.2%; Score 9; DB 2; Length 766;
 Best Local Similarity 100.0%; Pred. No. 6.2; Mismatches 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0;

QY 515 QELYAHGKH 523

Db 521 QELYAHGKH 529
 |||||

RESULT 6

Q9ZH42 PRELIMINARY; PRT; 93 AA.
 AC Q9ZH42; 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE Stability protein StbE.
 GN STBE.
 OS Morganella morganii (Proteus morganii).
 OG Plasmid R485.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Morganella.
 OX NCBI_TaxID=582;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99047595; PubMed=9829958;
 RA Hayes F.;
 RT "A family of stability determinants in pathogenic bacteria.";
 RL J. Bacteriol. 180:6415-6418 (1998).
 DR EMBL: AF072126; AAC84004.1; -
 KW Plasmid.
 SQ SEQUENCE 93 AA; 11024 MW; 373AB321776BDF2 CRC64;

Query Match 1.1%; Score 8; DB 2; Length 93;
 Best Local Similarity 100.0%; Pred. No. 10; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0;

QY 657 PRVPAARL 664

|||||


```

Db          37 PRVPAARL 44

RESULT 7
Q92EP9      PRELIMINARY;      PRT;      99 AA.

ID Q92EP9;
AC Q92EP9;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein lin0409.
DS LIN0409.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591975; CAC98470.1; -.
DR Listlist; LIN00409; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 99 AA; 11124 MW; 894D5703DCFE65B1 CRC64;

Query Match      1.1%; Score 8; DB 16; Length 99;
Best Local Similarity 100.0%; Pred. No. 11;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      2 AQTILKPI 9
Db          46 AQTILKPI 53

RESULT 8
Q8Y9X6      PRELIMINARY;      PRT;      99 AA.

ID Q8Y9X6;
AC Q8Y9X6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein lmo0391.
DS LMO0391.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591975; CAC98470.1; -.
DR Listlist; LIN00409; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 99 AA; 11124 MW; 894D5703DCFE65B1 CRC64;

Query Match      1.1%; Score 8; DB 16; Length 99;
Best Local Similarity 100.0%; Pred. No. 11;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      2 AQTILKPI 9
Db          46 AQTILKPI 53

RESULT 9
Q9DAA9      PRELIMINARY;      PRT;      171 AA.

ID Q9DAA9;
AC Q9DAA9;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE 1700016A09Rik protein.
DS 1700016A09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann M., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK006008; BAB24363.1; -.
DR MGD; MGI:1921477; 1700016A09Rik.
SQ SEQUENCE 171 AA; 19212 MW; 0D793A6487D6FFAC CRC64;

Query Match      1.1%; Score 8; DB 11; Length 171;
Best Local Similarity 100.0%; Pred. No. 18;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      479 AHRQTARS 486
Db          93 AHRQTARS 100

RESULT 10
Q9VNN6      PRELIMINARY;      PRT;      268 AA.

ID Q9VNN6;
AC Q9VNN6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

```

```

DE  CG1155 protein (LD24139P).
GN  SP558 OR CG1155.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=BERKELEY;
RC  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA  Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harjila N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA  Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Peinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
RA  Zheng X.H., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster."
RL  Science 287:2185-2195(2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Berkeley;
RA  Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA  Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA  Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA  Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA  Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL  Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AE003600; AAF51893.1; -.
DR  EMBL; AY058538; AAU13767.1; -.
DR  FlyBase; FBgn0040279; SP558.
SQ  SEQUENCE 268 AA; 2798 MW; 2A3B561F3B316BB3 CRC64;

Query Match 1.1%; Score 8; DB 5; Length 268;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 RSSNIELA 547
Db 67 RSSNIELA 74

RESULT 11

Query Match 1.1%; Score 8; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 VSGELGLR 179
Db 53 VSGELGLR 60

RESULT 13
Q93UWI

Q9NH68 PRELIMINARY; PRT; 268 AA.
ID Q9NH68
AC Q9NH68
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SP558 protein.
GN SP558 OR CG1155.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Serano T.L., Pendleton J.D., Rubin G.M.;
RT "A reverse genetic screen for genes involved in Drosophila eye
RT development";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF231039; AAF34808.1; -.
DR FlyBase; FBgn0040279; SP558.
SQ SEQUENCE 268 AA; 28078 MW; 414E78A9E9AE492 CRC64;

Query Match 1.1%; Score 8; DB 5; Length 268;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 540 RSSNIELA 547
Db 67 RSSNIELA 74

RESULT 12
Q9F7P9 PRELIMINARY; PRT; 398 AA.
ID Q9F7P9
AC Q9F7P9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Predicted NAD-dependent formate dehydrogenase.
OS uncultured proteobacterium EBAC31A08.
OC Bacteria; Proteobacteria; environmental samples.
OX NCBI_TaxID=133804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20446260; PubMed=10988064;
RA Beja O., Aravind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P.,
RA Jovanovich S.B., Gates C.M., Feldman R.A., Spudich J.L., Spudich E.N.,
RA DeLong E.F.;
RT "Bacterial rhodopsin: evidence for a new type of phototrophy in the
RT Sea";
RL Science 289:1902-1906(2000).
DR EMBL; AF299106; AAG10470.1; -.
DR HSSP; P33160; 2NAD.
DR InterPro; IPR006139; 2-Hacid_DH.
DR InterPro; IPR006140; 2-Hacid_DH_C.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR PROSITE; PS00065; D_2_HYDROXYACID_DH; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
SQ SEQUENCE 398 AA; 43992 MW; 7C2EC76FD5723C4C CRC64;

Query Match 1.1%; Score 8; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 VSGELGLR 179
Db 53 VSGELGLR 60

RESULT 13
Q93UWI

```

```

ID Q93UW1 PRELIMINARY; PRT; 399 AA.
AC Q93UW1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NAD+ dependent formate dehydrogenase.
GN FDH.
OS Hyphomicrobium sp. JCI17.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Hyphomicrobiales; Hyphomicrobium.
OX NCBI_TaxID=142666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCI17;
RA Tanaka Y., Yoshida T., Watanabe K., Mitsunaga T.;
RT "Characterization, gene cloning and expression of NAD+-dependent
RT formate dehydrogenase from a methylophilic bacterium Hyphomicrobium
RT sp. JCI17."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051073; BAB55449.1; -.
DR InterPro; IPR006139; 2-Hacid_DH.
DR InterPro; IPR006140; 2-Hacid_DH_C.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
SQ SEQUENCE 399 AA; 44044 MW; 853BAD2C635864DC CRC64;

Query Match 1.1%; Score 8; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
Db 54 VSGELGLR 61

RESULT 14
Q17898 PRELIMINARY; PRT; 399 AA.
ID Q17898;
AC Q17898;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE C10C5.3 protein.
GN C10C5.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews P.;
RC Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018 (1998).
DR EMBL; Z68214; CAA92445.1; -.
DR MEROPS; M20.973; -.
DR WormPep; C10C5.3; CE03011.
DR InterPro; IPR001261; ARGE_DAPE_CPG2.
DR InterPro; IPR002933; Peptidase_M20.
DR Pfam; PF01546; Peptidase_M20; 1.
DR PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
SQ SEQUENCE 399 AA; 45679 MW; 8036854C8D0A6D8B CRC64;

Query Match 1.1%; Score 8; DB 5; Length 399;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 15 LINTPLLA 22
Db 361 LINTPLLA 368

RESULT 15
Q9RSV0 PRELIMINARY; PRT; 400 AA.
ID Q9RSV0;
AC Q9RSV0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE NAD(+) dependent formate dehydrogenase (EC 1.2.1.2) (Fragment).
OS Mycobacterium vaccae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1810;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96158475; PubMed=8597552;
RA Galkin A., Kulakova L., Tishkov V., Esaki N., Soda K.;
RT "Cloning of formate dehydrogenase gene from a methanol-utilizing
RT bacterium Mycobacterium vaccae N10."
RL Appl. Microbiol. Biotechnol. 44:479-483 (1995).
DR HSSP; P33160; 2NAD.
DR InterPro; IPR006139; 2-Hacid_DH.
DR InterPro; IPR006140; 2-Hacid_DH_C.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
SQ SEQUENCE 400 AA; 44018 MW; 44228E88CA22BD9B CRC64;

Query Match 1.1%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
Db 53 VSGELGLR 60

RESULT 16
Q93GW3 PRELIMINARY; PRT; 400 AA.
ID Q93GW3;
AC Q93GW3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE NAD-dependent formate dehydrogenase.
GN FDH.
OS Paracoccus sp. 12-A.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Paracoccus.
OX NCBI_TaxID=171192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12-A;
RA Shinoda T., Satoh T., Mineki S., Iida M., Taguchi H.;
RT "Cloning, nucleotide sequence and expression in Escherichia coli of
RT the gene for formate dehydrogenase of Paracoccus sp. 12-A, a formate-
RT assimilating bacterium."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB071373; BAB64941.1; -.
DR InterPro; IPR006139; 2-Hacid_DH.
DR InterPro; IPR006140; 2-Hacid_DH_C.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF02826; 2-Hacid_DH_C; 1.
DR PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
SQ SEQUENCE 400 AA; 43757 MW; 13195AED9AC299D8 CRC64;

```

Query Match		1.1%; Score 8; DB 2; Length 400;
Best Local Similarity		100.0%; Pred. No. 39;
Matches		8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY		172 VSGELGLR 179
Db		54 VSGELGLR 61
RESULT 17		
Q93GV1		PRELIMINARY; PRT; 401 AA.
AC		Q93GV1;
DT		01-DEC-2001 (TRENBLrel. 19, Created)
DT		01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT		01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE		Formate dehydrogenase (EC 1.2.1.2).
GN		FDH.
OS		Mycobacterium vaccae.
OC		Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC		Cornebacterineae; Mycobacteriaceae; Mycobacterium.
OX		NCBI_TaxID=1810;
RN		[1]
RP		SEQUENCE FROM N.A.
RX		MEDLINE=96158475; PubMed=8597552;
RA		Galkin A., Kulakova L., Tishkov V., Esaki N., Soda K.;
RL		Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN		[2]
RP		SEQUENCE FROM N.A.
RX		MEDLINE=96158475; PubMed=8597552;
RA		Galkin A., Kulakova L., Tishkov V., Esaki N., Soda K.;
RT		"Cloning of formate dehydrogenase gene from a methanol-utilizing
RL		bacterium Mycobacterium vaccae N10.";
RT		Appl. Microbiol. Biotechnol. 44:479-483 (1995).
DE		EMBL; AB072394; BAB69476.1; -.
DR		InterPro; IPR006139; 2-Hacid_DH.
DR		InterPro; IPR006140; 2-Hacid_DH_C.
DR		Pfam; PF00389; 2-Hacid_DH; 1.
DR		Pfam; PF02826; 2-Hacid_DH; 1.
DR		PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.
DR		PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
DR		PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
KW		Oxidoreductase.
SQ		SEQUENCE 401 AA; 44149 MW; 8C4821B4790AAD26 CRC64;
Query Match		1.1%; Score 8; DB 2; Length 401;
Best Local Similarity		100.0%; Pred. No. 39;
Matches		8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY		172 VSGELGLR 179
Db		54 VSGELGLR 61
RESULT 18		
Q930E7		PRELIMINARY; PRT; 401 AA.
AC		Q930E7;
DT		01-DEC-2001 (TRENBLrel. 19, Created)
DT		01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT		01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE		Probable NAD-dependent formate dehydrogenase.
GN		RA0251 OR SWA0478.
OS		Rhizobium meliloti (sinorhizobium meliloti).
OC		Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC		Rhizobiaceae; Sinorhizobium.
OX		NCBI_TaxID=382;
RN		[1]
RP		SEQUENCE FROM N.A.
RC		STRAIN=1021;
RX		MEDLINE=21396509; PubMed=11481432;
RA		Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,		
Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,		
Kalm S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,		
Yeh K.-C., Davis R.W., Federapiel N.A., Long S.R.;		
"Nucleotide sequence and predicted functions of the entire		
RT		Sinorhizobium meliloti pSyma megaplasmid.";
RL		Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).
EMBL; AE007218; AAK64909.1; -.		
DR		InterPro; IPR006139; 2-Hacid_DH.
DR		InterPro; IPR006140; 2-Hacid_DH_C.
DR		Pfam; PF00389; 2-Hacid_DH; 1.
DR		Pfam; PF02826; 2-Hacid_DH; 1.
DR		PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.
DR		PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
DR		PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
KW		Plasmid; Complete proteome.
SQ		SEQUENCE 401 AA; 43533 MW; C2693A417456204B CRC64;
Query Match		1.1%; Score 8; DB 16; Length 401;
Best Local Similarity		100.0%; Pred. No. 39;
Matches		8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY		172 VSGELGLR 179
Db		56 VSGELGLR 63
RESULT 19		
O08375		PRELIMINARY; PRT; 402 AA.
ID		O08375;
DT		01-JUL-1997 (TRENBLrel. 04, Created)
DT		01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT		01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE		NAD-dependent formate dehydrogenase (EC 1.2.1.2).
GN		FDH.
OS		Moraxella sp.
OC		Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC		Moraxellaceae; Moraxella.
OX		NCBI_TaxID=479;
RN		[1]
RP		SEQUENCE FROM N.A.
RC		STRAIN=C-1;
RA		Galkin A.G.;
RL		Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
EMBL; Y13245; CAA73696.1; -.		
DR		HSP; E33160; 2NAD.
DR		InterPro; IPR006139; 2-Hacid_DH.
DR		InterPro; IPR006140; 2-Hacid_DH_C.
DR		Pfam; PF00389; 2-Hacid_DH; 1.
DR		Pfam; PF02826; 2-Hacid_DH; 1.
DR		PROSITE; PS00065; D_2_HYDROXYACID_DH_1; 1.
DR		PROSITE; PS00670; D_2_HYDROXYACID_DH_2; 1.
DR		PROSITE; PS00671; D_2_HYDROXYACID_DH_3; 1.
KW		NAD; Oxidoreductase.
SQ		SEQUENCE 402 AA; 44143 MW; C511CE1CB996BBB CRC64;
Query Match		1.1%; Score 8; DB 2; Length 402;
Best Local Similarity		100.0%; Pred. No. 39;
Matches		8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY		172 VSGELGLR 179
Db		54 VSGELGLR 61
RESULT 20		
Q9F771		PRELIMINARY; PRT; 427 AA.
ID		Q9F771;
DT		01-MAR-2001 (TRENBLrel. 16, Created)
DT		01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT		01-JUN-2001 (TRENBLrel. 17, Last annotation update)

```
DE Xerc.
GN XERC.
OS Pseudomonas aeruginosa.
OC Plasmid pKLC102.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SG17M;
RX MEDLINE=20476411; PubMed=11021913;
RA Kiewitz C., Larbig K., Klockgether J., Weinel C., Tummler B.;
RT "Monitoring genome evolution ex vivo: reversible chromosomal
RT integration of a 106 kb plasmid at two tRNA(Lys) gene loci in
RT sequential Pseudomonas aeruginosa airway isolates.";
RL Microbiology 146:2365-2373(2000).
DR EMBL; AF285416; AA02084.1; -.
DR InterPro; IPR002104; Phage_integrase.
DR Pfam; PF00589; Phage_integrase; 1.
KW Plasmid.
SQ SEQUENCE 427 AA; 48137 MW; 830637E70A0420A7 CRC64;

Query Match 1.1%; Score 8; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 HRAVLQWR 245
DB 46 HRAVLQWR 53

RESULT 21
P74655 PRELIMINARY; PRT; 544 AA.
ID P74655;
AC P74655;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein sll1550.
GN SLL1550.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90917; BAA18771.1; -.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 1.
DR PROSITE; PS01072; SLH_DOMAIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 544 AA; 58631 MW; 48A4905CA184F61A CRC64;

Query Match 1.1%; Score 8; DB 16; Length 544;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 RLSSGNLE 186
DB 222 RLSSGNLE 229

RESULT 22
Q9CMR7 PRELIMINARY; PRT; 925 AA.
ID Q9CMR7;
AC Q9CMR7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PM0745.
GN PM0745.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AS006112; AA02829.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 925 AA; 105081 MW; F46A9F52B59145E3 CRC64;

Query Match 1.1%; Score 8; DB 16; Length 925;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 GDYVRGRL 630
DB 696 GDYVRGRL 703

RESULT 23
O74360 PRELIMINARY; PRT; 933 AA.
ID O74360;
AC O74360;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE GTPase-activator protein for Rho-like GTPases.
GN SPBC28E12.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
DR EMBL; AL031517; CAA20650.2; -.
DR GenesDB Spombe; SPEC28E12.03; -.
DR InterPro; IPR001781; LIM.
DR InterPro; IPR00198; RhoGAP.
DR Pfam; PF00412; LIM; 2.
DR Pfam; PF00620; RhoGAP; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00132; LIM; 2.
DR SMART; SM00324; RhoGAP; 1.
DR PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
KW LIM domain; Metal-binding; Zinc.
SQ SEQUENCE 933 AA; 104487 MW; 2EE75150F62B5586 CRC64;

Query Match 1.1%; Score 8; DB 3; Length 933;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 EGLYRKSG 213
```

```

Db          779 EGLYRKSG 786
|||||
RESULT 24
Q8PD15      PRELIMINARY;      PRT;      984 AA.
ID  Q8PD15
AC  Q8PD15;
DT  01-OCT-2002 (TrEMBLrel. 22, Created)
DT  01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  Hypothetical protein XCC0531.
GN  XCC0531.
OS  Xanthomonas campestris (pv. campestris).
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC  Xanthomonadaceae; Xanthomonas.
OX  NCBI_TaxID=340;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 33913 / NCPPB 528;
RX  MEDLINE=22022145; PubMed=12024217;
RA  da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA  Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA  Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA  Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA  Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA  Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA  Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA  Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA  Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA  Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA  Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA  Pereira L.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA  Spinola H.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,
RA  Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
RA  Stubal J.C., Kitajima J.P.;
RT  "Comparison of the genomes of two Xanthomonas pathogens with differing
RT  host specificities."
RL  Nature 417:459-463(2002).
DR  EMBL; AE012151; AAM39847.1; -.
DR  InterPro; IPR000531; TonB_boxC.
DR  Pfam; PF00593; TonB_dep_Rec; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 984 AA; 108243 MW; 07603CAB01E0DC77 CRC64;

Query Match      1.1%; Score 8; DB 16; Length 984;
Best Local Similarity 100.0%; Pred.No.90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          747 GRSGTGGV 754
Db          703 GRSGTGGV 710
|||||
RESULT 25
Q06407      PRELIMINARY;      PRT;      1009 AA.
ID  Q06407
AC  Q06407;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE  Similar in C-terminus to chimaerin.
GN  RGA2 OR D9481.4 OR YDR379W.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX  NCBI_TaxID=4932;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=S288C;
RA  Ding H.;
RL  Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[2]

SEQUENCE FROM N.A.
STRAIN=S288C;
RA  Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA  Pavello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA  Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA  Johnston L., Langston Y., Latreille P., Le T., Mardis E., Meneses S.,
RA  Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA  Taich A., Trevisan E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA  Wilson R., Waterston R.;
RL  Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=S288C;
RA  Waterston R.;
RL  Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[4]
RP  SEQUENCE FROM N.A.
RC  STRAIN=S288C;
RA  Jia Y., Cherry J.M.;
RL  Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR  EMBL; U28373; AAB64815.1; -.
DR  HSSP; P27986; 1PBW.
DR  SGD; S0002787; RGA2.
DR  InterPro; IPR001781; LIM.
DR  InterPro; IPR001198; RhoGAP.
DR  Pfam; PF00412; LIM; 2.
DR  Pfam; PF00620; RhoGAP; 1.
DR  ProDom; PD000094; LIM; 2.
DR  SMART; SM00132; LIM; 2.
DR  SMART; SM00324; RhoGAP; 1.
DR  PROSITE; PS00478; LIM DOMAIN 1; 1.
KW  LIM domain; Metal-binding; Zinc.
SQ  SEQUENCE 1009 AA; 113290 MW; FA13BALDF8B35D98 CRC64;

Query Match      1.1%; Score 8; DB 3; Length 1009;
Best Local Similarity 100.0%; Pred.No.92;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          206 EGLYRKSG 213
Db          822 EGLYRKSG 829
|||||
RESULT 26
Q8H614      PRELIMINARY;      PRT;      2319 AA.
ID  Q8H614
AC  Q8H614;
DT  01-MAR-2003 (TrEMBLrel. 23, Created)
DT  01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  Putative gag-pol polyprotein.
OS  Zea mays (Maize).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC  PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX  NCBI_TaxID=4577;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CV. BSSS53;
RX  MEDLINE=22253764; PubMed=12368247;
RA  Song R., Liaca V., Messing J.;
RT  "Mosaic Organization of Orthologous Sequences in Grass Genomes.";
RL  Genome Res. 12:1549-1555(2002).
DR  EMBL; AF528565; AAN40030.1; -.
KW  Polyprotein.
SQ  SEQUENCE 2319 AA; 263271 MW; 56378EC6E16EDF08 CRC64;

Query Match      1.1%; Score 8; DB 10; Length 2319;
Best Local Similarity 100.0%; Pred.No.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          92 ASAPVIRG 99
|||||

```

Db 1807 ASAPVIRG 1814

RESULT 27

Q8GBX6 PRELIMINARY; PRT; 5017 AA.

ID Q8GBX6 AC Q8GBX6

DT 01-MAR-2003 (TReMBLrel. 23, Created)

DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Polypeptide synthase.

GN SPIH.

OS Polyangium cellulosum.

OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;

OC Sorangineae; Polyangiaceae; Polyangium.

OX NCBI_TaxID=56;

RN [1]

RP SEQUENCE FROM N.A.

RA Knauber J., Bloeker H., Mueller R., Nordieck G., Beyer S.;

RT "The spirangine synthase from Sorangium cellulosum So ce90 - Module

RT duplications as strategy behind the evolution of polyketide

RT synthases.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ505006; CAD43449.1; -

SQ SEQUENCE 5017 AA; 525555 MW; 08BED56182669493 CRC64;

Query Match 1.1%; Score 8; DB 2; Length 5017;

Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 GLVDVADG 161

Db 2703 GLVDVADG 2710

RESULT 28

Q8KY74 PRELIMINARY; PRT; 48 AA.

ID Q8KY74 AC Q8KY74

DT 01-OCT-2002 (TReMBLrel. 22, Created)

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Alpha-crystallin protein (Fragment).

GN ACR.

OS Mycobacterium habana.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1784;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=M7052;

RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaels F.;

RT "Detection of a gene encoding an alpha-crystallin homolog in numerous

RT mycobacterial species other than Mycobacterium tuberculosis.";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF253458; AAM6920.1; -

DR InterPro; IPR002068; Hsp20.

DR Pfam; PF00011; HSP20; 1.

FT NON TER 1

SQ SEQUENCE 48 AA; 5491 MW; 14DD8B8694296C71 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 48;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340

Db 23 RYELRAE 29

RESULT 29

Q8KY83 PRELIMINARY; PRT; 48 AA.

ID Q8KY83 AC Q8KY83

DT 01-OCT-2002 (TReMBLrel. 22, Created)

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Alpha-crystallin protein (Fragment).

GN ACR.

OS Mycobacterium triviale.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1798;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AFCC23292;

RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaels F.;

RT "Detection of a gene encoding an alpha-crystallin homolog in numerous

RT mycobacterial species other than Mycobacterium tuberculosis.";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF253446; AAM69208.1; -

DR InterPro; IPR002068; Hsp20.

DR Pfam; PF00011; HSP20; 1.

FT NON TER 1

SQ SEQUENCE 48 AA; 5477 MW; 7FB68B8694296820 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 48;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340

Db 23 RYELRAE 29

RESULT 30

Q8KHQ1 PRELIMINARY; PRT; 48 AA.

ID Q8KHQ1 AC Q8KHQ1

DT 01-OCT-2002 (TReMBLrel. 22, Created)

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Alpha-crystallin protein (Fragment).

GN ACR.

OS Mycobacterium ulcerans.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1809;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC19423, Congo, and Mexico;

RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaels F.;

RT "Detection of a gene encoding an alpha-crystallin homolog in numerous

RT mycobacterial species other than Mycobacterium tuberculosis.";

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF253450; AAM69212.1; -

DR EMBL; AF253451; AAM69213.1; -

DR EMBL; AF253452; AAM69214.1; -

DR InterPro; IPR002068; Hsp20.

DR Pfam; PF00011; HSP20; 1.

FT NON TER 1

SQ SEQUENCE 48 AA; 5477 MW; 7FB68B8694296820 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 48;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340

Db 23 RYELRAE 29

RESULT 31

Q08482 PRELIMINARY; PRT; 55 AA.

ID Q08482 AC Q08482

```

DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-FEB-1997 (TrEMBLrel. 02, Last annotation update)
DE L protein (Fragment).
GN L.
OS Viral hemorrhagic septicemia virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Rhabdoviridae; Novirhabdovirus.
OX NCBI_TaxID=11287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FI13;
RA Schuetze H.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X73873; CAA52082.1; -.
FT NON TER 55
SQ SEQUENCE 55 AA; 6352 MW; 934733D8EAX24A1E CRC64;

Query Match 0.9%; Score 7; DB 12; Length 55;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 HOERLPS 513
DB 11 HOERLPS 17

RESULT 32
Q8L411 PRELIMINARY; PRT; 60 AA.
AC Q8L411;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE QJ1103.E04.11 protein (P0675B10.1 protein).
GN QJ1103.E04.11 OR P0675B10.1.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T.; Matsumoto T.; Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
clone:OJ1103.E04.11";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T.; Matsumoto T.; Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
clone:P0675B10.1";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003806; BAC06954.1; -.
DR EMBL; AF004347; BAC07130.1; -.
DR Gramene; Q8L411; -.
SQ SEQUENCE 60 AA; 6630 MW; C799493C3836E719 CRC64;

Query Match 0.9%; Score 7; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 RLPDSR 231
DB 12 RLPDSR 18

RESULT 33
Q9JSV8 PRELIMINARY; PRT; 76 AA.
AC Q9JSV8;

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein NMA2114.
GN NMA2114.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RA MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491."
RL Nature 404:502-506(2000).
DR EMBL; AL162758; CAB85328.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 76 AA; 8826 MW; 9DB776C1C687DB48 CRC64;

Query Match 0.9%; Score 7; DB 16; Length 76;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 674 DRIDANL 680
DB 17 DRIDANL 23

RESULT 34
Q8KY84 PRELIMINARY; PRT; 86 AA.
AC Q8KY84;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium nonchromogenicum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19530;
RA Shampura I.C., Bastian I., Fonteyne P.-A., Portaeis P.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
mycobacterial species other than Mycobacterium tuberculosis."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253445; AAM69207.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON TER 86
FT NON TER 86
SQ SEQUENCE 86 AA; 9722 MW; BAF085B56BF1DBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

RESULT 35

```



```
Q8KY71
ID Q8KY71 PRELIMINARY; PRT; 86 AA.
AC Q8KY71
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium gastris.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M97-938;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253461; AAM69223.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9722 MW; BAF085B56BFD1DBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

RESULT 36
Q8KY78
ID Q8KY78 PRELIMINARY; PRT; 86 AA.
AC Q8KY78
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Japan;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253454; AAM69216.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9695 MW; A5D2B8AA472C35F3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

Q8KY77
ID Q8KY77 PRELIMINARY; PRT; 86 AA.
AC Q8KY77
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium marinum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC927;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253449; AAM69211.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9694 MW; A5D2B8AA49CCDBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

RESULT 38
Q8KY79
ID Q8KY79 PRELIMINARY; PRT; 86 AA.
AC Q8KY79
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=China;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253453; AAM69215.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9694 MW; A5D2B8AA49CCDBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29
```

```
RESULT 37
Q8KY80
ID Q8KY80 PRELIMINARY; PRT; 86 AA.
AC Q8KY80
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium marinum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC927;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253449; AAM69211.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9694 MW; A5D2B8AA49CCDBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29

RESULT 38
Q8KY79
ID Q8KY79 PRELIMINARY; PRT; 86 AA.
AC Q8KY79
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=China;
RA Shamputa I.C., Bastian I., Fonteyne P.-A., Portaeels F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RT mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253453; AAM69215.1; -.
DR InterPro; IPR002068; Hsp20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON_TER 1
FT NON_TER 86
SQ SEQUENCE 86 AA; 9694 MW; A5D2B8AA49CCDBF3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29
```

```
RESULT 39
Q8KY77
ID Q8KY77 PRELIMINARY; PRT; 86 AA.
AC Q8KY77;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Alpha-crystallin protein (Fragment).
GN ACR.
OS Mycobacterium ulcerans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1809;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Surinam;
RA Shamputa I.C., Bastian I., Ponteyne P.-A., Portaela F.;
RT "Detection of a gene encoding an alpha-crystallin homolog in numerous
RL mycobacterial species other than Mycobacterium tuberculosis.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF253455; AAM69217.1; -.
DR InterPro: IPR002068; HSP20.
DR Pfam: PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
FT NON TER 1
FT NON TER 86
SQ SEQUENCE 86 AA; 9778 MW; A5D2B8B09DEC35F3 CRC64;

Query Match 0.9%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 RYELRAE 340
DB 23 RYELRAE 29
|||||

RESULT 40
Q9BK93
ID Q9BK93 PRELIMINARY; PRT; 91 AA.
AC Q9BK93;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Polythreonine-rich glycoprotein (Fragment).
OS Cryptosporidium sp. 'muskrat 603'.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=138541;
RN [1]
RP SEQUENCE FROM N.A.
RC NEDLINE=21141828; PubMed=11229905;
RA Perz J.F., Le Blancq S.M.;
RT "Cryptosporidium parvum Infection Involving Novel Genotypes in
RL Wildlife from Lower New York State.";
RL Appl. Environ. Microbiol. 67:1154-1162(2001).
DR EMBL; AF303055; AAK20279.1; -.
FT NON TER 1
FT NON TER 91
SQ SEQUENCE 91 AA; 10009 MW; 794C7FEEDAC9D19C CRC64;

Query Match 0.9%; Score 7; DB 5; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTLKPIV 10
DB 24 TTLKPIV 30
|||||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:57:37 ; Search time 22 seconds
(without alignments)
1457.801 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLPKPVLSILLINPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 4231058 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2_6/prodata/1/iaa/5A COMB.pap.*
- 2: /cgm2_6/prodata/1/iaa/5B COMB.pap.*
- 3: /cgm2_6/prodata/1/iaa/6A COMB.pap.*
- 4: /cgm2_6/prodata/1/iaa/6B COMB.pap.*
- 5: /cgm2_6/prodata/1/iaa/PTUS COMB.pap.*
- 6: /cgm2_6/prodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1391	34.5	704	4	US-09-328-352-5144
2	685	17.0	718	4	US-09-252-991A-29432
3	306	7.6	741	4	US-09-328-352-5898
4	246	6.1	753	4	US-09-252-991A-17987
5	232.5	5.8	789	4	US-09-252-991A-27011
6	227	5.6	652	1	US-08-765-081-6
7	227	5.6	652	3	US-09-098-082-6
8	216	5.4	713	3	US-08-628-434-4
9	203	5.0	725	3	US-08-425-843-2
10	202	5.0	713	3	US-08-628-434-2
11	198	4.9	617	4	US-09-252-991A-30921
12	193	4.8	707	4	US-09-328-352-4834
13	189	4.7	715	3	US-08-425-843-7
14	188.5	4.7	663	1	US-08-765-081-7
15	188.5	4.7	663	3	US-09-098-082-7
16	182	4.5	718	5	PCR-US95-06994-7
17	180.5	4.5	914	1	US-08-487-890A-11
18	180.5	4.5	914	2	US-08-478-435-11
19	180.5	4.5	914	2	US-08-337-483-11
20	180.5	4.5	914	2	US-08-478-373-11
21	180.5	4.5	914	3	US-08-474-671-11
22	180.5	4.5	914	3	US-08-483-577A-11
23	180.5	4.5	914	3	US-08-897-438-11
24	180.5	4.5	914	4	US-08-637-654-11
25	180.5	4.5	914	4	US-08-649-518-11
26	179	4.4	756	4	US-09-252-991A-30809
27	174	4.3	761	4	US-09-328-352-5942

28	173	4.3	723	4	US-09-252-991A-18279	Sequence 18279, A
29	170.5	4.2	706	4	US-09-252-991A-32448	Sequence 32448, A
30	169	4.2	595	4	US-09-252-991A-25815	Sequence 25815, A
31	165.5	4.1	1245	4	US-09-252-991A-30935	Sequence 30935, A
32	164.5	4.1	643	4	US-09-328-352-5146	Sequence 5146, Ap
33	163	4.0	913	1	US-08-487-890A-5	Sequence 5, Appl
34	163	4.0	913	2	US-08-478-435-5	Sequence 5, Appl
35	163	4.0	913	2	US-08-337-483-5	Sequence 5, Appl
36	163	4.0	913	2	US-08-474-671-5	Sequence 5, Appl
37	163	4.0	913	3	US-08-483-577A-5	Sequence 5, Appl
38	163	4.0	913	3	US-08-897-438-5	Sequence 5, Appl
39	163	4.0	913	4	US-08-637-654-5	Sequence 5, Appl
40	163	4.0	913	4	US-08-649-518-5	Sequence 5, Appl
41	163	4.0	912	1	US-08-487-890A-7	Sequence 7, Appl
42	161.5	4.0	912	1	US-08-487-890A-9	Sequence 9, Appl
43	161.5	4.0	912	2	US-08-478-435-7	Sequence 7, Appl
44	161.5	4.0	912	2	US-08-478-435-9	Sequence 9, Appl
45	161.5	4.0	912	2	US-08-478-435-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1

US-09-328-352-5144
; Sequence 5144, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5144
; LENGTH: 704
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5144

Query Match 34.5%; Score 1391; DB 4; Length 704;

Best Local Similarity 40.6%; Pred. No. 5.2e-125; Indels 118; Gaps 24;
Matches 315; Conservative 121; Mismatches 222;

Qy	9	IVLSIL-LINTPLLAQAHETEVSGLTETVTVGKSRPRATSGLLHTS---TASDKTISGD 64
Db	21	LTLSILAAISIPVFAAENVEK--LETTIRI--KAHP-----LEQTSKDFAVADTVDDQK 71
Qy	65	TLRQKAVNLGDALDGVPGIHASOYGGASAPVIRGOTGRRIKVLNHHGETGDMADFSPDH 124
Db	72	HLTEGATIGDALNSEVGIYANQFAGSRFPVIRGQDGPVKVLQNSSENVVSTLSPDH 131
Qy	125	AIWVDTALSOQVRIIRGPVTLTYSSNGVAGLVADGKIPKMPENGVSGLRLSSGN 184
Db	132	AVTVDPVLAQVEVIRGPTLLFGAGTVGLVNVINDNKIPTQMPENGVEQVGLRYNTGS 191
Qy	185	LEKLTSGINIGLGNFVLHTEGLYKSGDYAPRY-----RNLRKLPDSPRRANGQHRA 240
Db	192	DEKLASAGTVGLGSQVALRVEGLTDANNYIAPNTIHEGEKERRVDNT---FAQGDSDVN 248
Qy	241	V-LGWRKRFYRR-----TVSDRRDQGLPAHSHEYDDCHADIWQKSLINKRYLYPHL 294
Db	249	VGLSW---IYDRGVYTGISYNSREDQYGLPCHSHEYSCSA-----HL 287
Qy	295	LTREDVDYDNPGLSCGFHDDD-----DAHAHA-----HNGKPTDLNRYELRAEWKQPF 345
Db	288	-----GGRPHLHCAHEHDEHEGEEHAHEHEHGGPWIDLKSEYDFKTELNDPF 340
Qy	346	PGFEALRVHLNRNDYHHDKAGDAVENFNQONARIELRHOPIGRLKSGMWQVYLGOK 405
Db	341	AGFQKLRAQASYDYQHDEIEEGTITATRFQNGYDRIELVHNPIADWEGVIGTQ-LGQQ 399

```

Query Match      7.6%; Score 306; DB 4; Length 741;
Best Local Similarity 20.5%; Pred. No. 2.5e-20;
Matches 163; Conservative 125; Mismatches 337; Indels 172; Gaps 32;

Qy 59 KIISGDTLRQAVNLGDALDQVGIHASQYGGCASAPVIRGTGRRIKVLNHHGTGMA 118
      ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 16 KYESSNT-----IGDALKNITCIQSTSFGNAGAPVIRLSGNRVGVIENGFMN 68

Qy 119 DRSPDHAIWDTALSSQVEILRGVTVLLYSSGNVAGLDVADCKIPKMPEN----- 170
      ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 69 AFSGHINIFFPIEFKIVVNKNTDNIRYGGNAIGGSVQIESGLIPKKIEEKNKLDIVF 128

Qy 171 -----GVSG-----ELGLRLSS-----GNLE-KI-----TSGGINI 195
      ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 129 RKGFNDFDVGFNFNINDQKNWSTNIRYSVEIYSYKIPGNSKAKI-CEDQIFNSGGINS 188

Qy 196 GLGKNF-----VLHTEGLYKSGDYAVPRYENLKRLLPDSFRPFANGQRAVLGWRKYR 250
      ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 189 ALAASQKDSRVQH-----IYNKSSOPYIDKF--MTENPD-----WADGDF-----SYTF 231

Qy 251 -----RTYSRRDQYGLP-AHSHEYDDCHADII--WQSLNKRYQLQYPLHLLTEE 298
      ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db 232 DKPTSIWGKVIYNPKNPEIYNTPTNTKINTDVTNPFKLGNS-YAO-----NE 283

```


Db 323 STL---APIGHEMLSPGVE--GKHESLEDKTSNKSISRTHISN-TQWAGFIEDEWALAE 376
Qy 441 NFTLEGGVVRKQKASIRYDKALIDRENYKQPLDGLAHROTARSPALSGNWTTPQHK 500
Db 377 QFRLTFGRLDHDK-----GSH-----FSPRVYGVNLDPLMT 411
Qy 501 LSLTASHQERLPSTQELAHGKHVATNTFEVGNKHLNKNERSNNIELALGYE-GDRWQYNL 559
Db 412 VKGGVSTGFAPQLREVTPDWQVGGGNIYGNPDLPQETSINKELSLMYSTGSLAASL 471
Qy 560 ALYENRF-----GNVIAQTLNDGRGPKSIEDSEMMLVRYNOSGADFYGAEGEI 609
Db 472 TAFHNDFKDKITRVACPNICTAGPNQWATP-----TYRVNIDEAETYGAEATL 521
Qy 610 YFKPTPRYRIGVSGDYVRGRLKNLPSLPGRDAYSNGRPFIAQDDQNAVRPAARLGFHLK 669
Db 522 SLPTESVELSSSYTYTHSEQS-----GNFAGRELL-----QLPKHLFNANLS 565
Qy 670 ASLTDRIID--ANLDY--YRVFAQNKLYETRTFEGHMLNLGANYRRNTRYGENWYVKA 725
Db 566 WQTTDLNSWANLYRGKEMQPEGASNDDFIAPSYTFIDTGVYALTD-----TATIKA 620
Qy 726 D--NLLNOSV-YAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 621 AVYNLFDOEVNAYEYGYVED-----GRRYWLGLDIAF 652

RESULT 7

US-09-098-082-6
; Sequence 6, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli O157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrgA amino acid sequence
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:

; ORGANISM: Vibrio Cholerae
US-09-098-082-6

Query Match 5.6%; Score 227; DB 3; Length 652;
Best Local Similarity 21.6%; Pred No. 8.3e-13;
Matches 176; Conservative 108; Mismatches 310; Indels 222; Gaps 40;
Qy 1 MAQTLKPIVLSI-----LLINTPLLAQ-AHETEQSVGLETTVVVGVKSRPRATSGLLHTSTA 56
Db 1 MSRFNPSVSLSVTLMLFSAFAQDATKTDETM---VVTAAGYAQ-----VIQNAFA 51
Qy 57 SDKIISGTLIQKAV-NLGDALDGVPGIHASQYGGASAPV-IRG-----Q 100
Db 52 SISVISREDLESRYRVDYDALKSVPGVTVT--GGGDTTDSIRMGSGNYTLILVDGRKQ 109
Qy 101 TGRRIKVLNHGETGDMADFSFDHAIWDTALTALSOQVEILRGVPTLTYSSGNVAGLVADVAD 160
Db 110 TSQTRP-NSDGPGLIEQGLPPLQAI-----ERIEVIRGPMSTLYGSDALGGVINIT 161
Qy 161 GKIPKMPENGVSGBELGRLSSGNLEKLTSGGINIGLGNFVLHTEGLYKSGDYAVPRY 220
Db 162 RKDQOQW-----SGNVQLSTVVQEN-----RASGD----- 186
Qy 221 RMLKELPDSRRFANGQHRVGLGWRKRPYRTYSDRRQYGLPAHSHEYDDCHADLIWQK 280
Db 187 -----EOSANFFVTGPLSDALS--LQYVGQTTORDEDEI-----EHGYGDKSL-----R 228
Qy 281 SLINKRYLQLYP--HLLTEEDVD---YDNPGLS-----CGFHDDDDAHAH---AH 322
Db 229 SLTSKINYQLNPDHQLQLEAGVSAQDRENNVKSQSSGCRGCTCNTDNYRRNHVAVSH 288
Qy 323 NGKQWIDLNRKRYELRAEWKQFPFGFEALRYHLNRNDYH--HDEKAGDAVENFFNNQTN 380
Db 289 QG-----DWQ-----GVGQSDTYLQVEENTNKSREMSIDNTVPK 322
Qy 381 ARIELRHOPIGELKSGWQVYLQKSSALSATSEAVKQPMLLDNKVQHYSPFVQOANWD 440
Db 323 STL---APIGHEMLSPGVE--GKHESLEDKTSNKSISRTHISN-TQWAGFIEDEWALAE 376
Qy 441 NFTLEGGVVRKQKASIRYDKALIDRENYKQPLDGLAHROTARSPALSGNWTTPQHK 500
Db 377 QFRLTFGRLDHDK-----GSH-----FSPRVYGVNLDPLMT 411
Qy 501 LSLTASHQERLPSTQELAHGKHVATNTFEVGNKHLNKNERSNNIELALGYE-GDRWQYNL 559
Db 412 VKGGVSTGFAPQLREVTPDWQVGGGNIYGNPDLPQETSINKELSLMYSTGSLAASL 471
Qy 560 ALYENRF-----GNVIAQTLNDGRGPKSIEDSEMMLVRYNOSGADFYGAEGEI 609
Db 472 TAFHNDFKDKITRVACPNICTAGPNQWATP-----TYRVNIDEAETYGAEATL 521
Qy 610 YFKPTPRYRIGVSGDYVRGRLKNLPSLPGRDAYSNGRPFIAQDDQNAVRPAARLGFHLK 669
Db 522 SLPTESVELSSSYTYTHSEQS-----GNFAGRELL-----QLPKHLFNANLS 565
Qy 670 ASLTDRIID--ANLDY--YRVFAQNKLYETRTFEGHMLNLGANYRRNTRYGENWYVKA 725
Db 566 WQTTDLNSWANLYRGKEMQPEGASNDDFIAPSYTFIDTGVYALTD-----TATIKA 620
Qy 726 D--NLLNOSV-YAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 621 AVYNLFDOEVNAYEYGYVED-----GRRYWLGLDIAF 652

RESULT 8

US-08-628-434-4
; Sequence 4, Application US/08628434
; Patent No. 6265567
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Beucher, Margaret
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule and Vaccine
; NUMBER OF SEQUENCES: 4

Query Match 5.0%; Score 203; DB 3; Length 725;
Best Local Similarity 19.8%; Pred. No. 2e-10;


```
Matches 158; Conservative 134; Mismatches 300; Indels 206; Gaps 43;
QY 11 LSILLINTPLAQAHETEOSVGLTWTVVVKSRPRATSGLLHTSTASDKIISGDTLRQKA 70
Db 6 LSLAIAPTT--LVTANALAQSVELDSINVIATRDP---SREAYTPEKQSK--DSLISKQA 57
QY 71 VNLGDALDGVPGFIHASQYGGG---ASAPVIRGOTGR-IKVLNHHGETGMDWADSPDH-- 124
Db 58 TSVAALLEDIPNDVIR--GGSRSIAQAPNIRGLSDNRVQVIDGVRFDLA-----HRG 110
QY 125 AIMVDTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPEKMPENGVSGLGLRLSSG- 183
Db 111 SYELPMSLIQIEIVIKPSSSLGSGALGGVAMRTPNALDLLKNN--DKFGVKIRQGY 167
QY 184 -NLEKLTSGGINI-GLKGNFVLHTEGLYKSGDYAVPRYRNLRKLPDPSRPFANGQHRV 241
Db 168 QTANNLSERDVSVFAANDKFDVLISGFYNNADNLRTGKGNLNN--TAYKQFGG---LAK 222
QY 242 LGWRKRFYRRT--YSDRRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDV 300
Db 223 FGQINDANVELSHRETRFKQTAPSN-----NEVENELTNQGITDQIR 266
QY 301 DYNPGLSCGFHDDDDAHAAHAKPWIDLRNRYELRAEWKQFPFGFEALRYHLNRNDY 360
Db 267 EFHKP-----NNGSP-----PRAKPSQEEF-----YSGVKTRFGSVSY 299
QY 361 HHDEKAGD--AVENFR---NNQTONARIELRHQPIGRLKSGWGVYILGQKSSALSATSEA 415
Db 300 LTDQQLPDQSTVFNYITPDNPVINTHIALYNNKT-----IEKQKVSQVKDQTKL 351
QY 416 VKQPMLLDN--KVQHSYF-FGVQEQANWDFTLBGGV-----RVEKQKA----- 455
Db 352 TTRGINLRNSELSHISFVGVVDYMR-DKIRTERGTNNKDAQFRADPYNANSNTTGYILI 410
QY 456 -----STRYD-----KALIDRENYKQPLDLAGHRQTAARSFALSGNWYF 495
Db 411 AH1PLFGEKLLSPSVRYDHYDTSSTVKYKDNHL-----SPATKLTWIV 455
QY 496 TPOHKLSTASHOE--RLPSTQELYAHGKHVAT-----NTEFVGNKHLAKERNNIE 545
Db 456 T--NWLDFATAKNEAFAPSMQERFVSGSHFTSILGRNEINKEF-VANPNLRPETAKNKE 512
QY 546 L-----ALGYBGDRQWYNALALRYRNFNGYIYAQTLINDGRGPKSTEDDSEMKLVRYNOS 598
Db 513 ITANLHFDSLFGQDKFKIEATYFRNDVKDFINLKIFNDAKTNNASASA-----GA 564
QY 599 GADFYCAEGEIFYKPT-PRYR-----IGVSGDYVRGRL---KNLPSLPQREDAYGN 645
Db 565 GA---GANPENGALLPTKSYQNTNARLSGIELQAOYQTERLTLFTNYGSTKGKDKDSG- 620
QY 646 RPFIAQQDNAPRVPAPARLGFHLKASITD---RIDANLDYRVFAQNKLAHYETRT-PGH 701
Db 621 -----EALSNIASKIGVGVNYALVKDKFTVGATVTTHYA--AQRVPKDHSHVTPSY 670
QY 702 HMLNLCANRYRNTRYGEW 719
Db 671 ILTDLRAATYA--PLKGEW 686

RESULT 10
US-08-628-434-2
; Sequence 2, Application US/08628434
; Patent No. 6265567
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Beucher, Margaret
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule and Vaccine
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
```

```
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,434
; FILING DATE: 05-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/418,964
; FILING DATE: 07-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-3-P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-628-434-2
```

```
Query Match 5.0%; Score 202; DB 3; Length 713;
Best Local Similarity 21.8%; Pred. No. 2.5e-10;
Matches 187; Conservative 97; Mismatches 306; Indels 274; Gaps 46;
QY 11 LSILLINTPL-LAQAHETEOSVGLTWTVVVKSRPRATSGLLHTSTASDKIISGDTLRQK 69
Db 8 LSLLSITLAAGFAHAENNVANVALDIVTVKGRQ-----GSKTRTNIV--TLQOK 55
QY 70 ----AVNLGDALDGVPGFIHASQYGGASAPVI-----RGQTGRRIKV-----LNHH 111
Db 56 DESTATDMRELLKEEPSI---DFGGNGTSGQFLTRGMGQNSVDIKVDNAYSQSIILYHQ 112
QY 112 GETGDMADSPDHAIMVDTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPEKMP-- 169
Db 113 GR-----FIVDPALVKVSVQKG-----AGSAGIGATNGAIIAKTVDAQ 153
QY 170 ---NGVSGELGLRLSSG--NLEKLTSGGINIGLKNFVLHTEGLY-----RKSGDYAVPR- 219
Db 154 DLLKGLDKWGVRLNSGFAGNCGASYGASVFGKEGNF---DGLFSYNRNDEKDYAGKG 209
QY 220 YRN---LKLPLDSP-----RRFANGQHRVAVLGMWRKFRYR--RTYSDRRDQYGLP 263
Db 210 FRNDNGKTVPYSAIDKRSYLAKIGTTFGDDHRIVLSHMKDQHRGIRTV--REEFAVS 266
QY 264 AHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHN 323
Db 267 EKN-----SRITIKRQAPSRYE--TTQSNNTLAYTKGLGFVEKLDANAYV-- 310
QY 324 GRPWIDLRNRYELRAEWKQFPFGPEALRVHLNRNDYHHDEKAGDAVENFFNNQTNARI 383
Db 311 -----LEKKRYS--ADKDK-----NGYAGNVKGPNTRIATRSMMNFDS 348
QY 384 ELRHQPIGRLKSGWGVYILGQ--KSSAL-----SATSEAVKQPMLLDNKVQHSY 430
Db 349 RLAEQTL--LK--YGINYRHQBIKQAPLNSBFEIKDEKATNEEKKKRENEKIAKAYR 404
QY 431 FGVGEQANW-----DNFTLBGGVVEKQKASIRVDKALIDRENYKQPLDLAGH 480
Db 405 LTNPTKTDGTAYIEAIEHDEGFTLTGLGFRYDKVTKTHDGK----- 445
QY 481 RQTARSAFALSGNW--YFTPOHKLSTASHQ--ERLPSTOE-LYAHGKH-----VATNTPEV 531
Db 446 --TVSSSSLNPSFGVWQPREHWSFSASHNYAGRSPRLYDALQTHGKRGIIISADGT--- 500
```

Qy	532	GNKHLKERSNNIELALGYEGDRWOYNIALYRNFGNYIVAOTLNDG-RGPKIEIDSEM	590
Db	501	-----KAERARTEIGFYNDGTFAAN-----GSY-FRQTIKALANPQRHDSVAV	546
Qy	591	K-----LVRYNQSGADFYGAEGEIVFKPTPRVRIQVSGD	624
Db	547	REAVNAGYIKOHGYELGASRYTGGLTAKVGVSHSKPRFYDTHKDKLLSANPEGAQVGT	606
Qy	625	YVRG---RLKNIPLSIP-----GR--EDAYGNRPFFIAQDDONAPRVPAARLGHFKASLTDR	675
Db	607	WTASLAYREPKN-ENLEIGWRGRYVQKAVGSLAAGQKDRGLKNVVVRQGFVN-----	659
Qy	676	IDANLDYVFEAQNKLARVERETPGHHMLNLGANYRNRTRYGEWNWVVKADNLINLSVYA	735
Db	660	-----DVFANWPKLGKDTLN-----VNLNVN-----NVFDEKYYP	689
Qy	736	HSSFLSDT-PQMGSRFTGGVNVKF	758
Db	690	HSORWTNTILPGVGRDVRLLGVNVKF	713

```

RESULT 11
US-09-252-991A-30921
; Sequence 30921, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30921
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30921

```

Query Match	4.9%;	Score 198;	DB 4;	Length 617;
Best Local Similarity	22.4%;	Pred. No. 4.7e-10;		
Matches 155;	Conservative 88;	Mismatches 249;	Indels 199;	Gaps 39;
QY	38	VVGKSRPRATSGLLHTSTASDKIISGDTL-RQKAVNLGDALDVGPIHASQYGG-GASAP	95	
Db	32	VVTATRTAQITAS--QSLAAVSVIDREDIERSQARSVPELLRQVPGVSLANNGYFGKNTT	88	
QY	96	V-IRGQTGRRIKVLNNHGETG-MADFPDPHAIMVDTSALSOQVEILRGFVTLTSSGNGVA	153	
Db	89	LFLRTGSDHVLVLIDGIKVGSASAGLTAFOQLPVE--LIERIEVVRGPRSSLVGEAIG	146	
QY	154	GLAVDY----ADGKIPEKMPENG-----VSGELGRLSSGNLEKLTSGGINIGLGNFVL	203	
Db	147	GVIQIFTRRGDQGAKPFFSAGYGTHTLEGSAGVSGAGN-----GWYSLGVSS---F	197	
QY	204	HTEGILYRK---SGDYAVPR--YRNKLRLPDSPPRFANGCHRAVLGRKFRYRTYSDDRD	258	
Db	198	DTAGINTKRAFTAGYEPDRDGYRNLSGNLRGGRYDFDNLG-----ELD	239	
QY	259	OYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDDAH	318	
Db	240	GTLLRKAGHN-----DYOQVFGNSGFNANADGE	267	
QY	319	AHAHNGK-----PW-IDLRNRYELRAEMKQP--PFGFEALRVHLN-RNDYHDEKA	366	
Db	268	QNLVGRARFTPFDPWLVTLOAGRSDKADAYQDGRFYSRFDTTRDSLSQWDLTLAB--	325	
QY	367	GDAVENFFNNTQNARIELRHQPIGLKGSWGVOYLQKQSSALSATSEAVKQPMLLDNKY	426	

```

326  GHVLTILGYDWQKDETSSEAFVSDBRNKNGFWAQLQGYG-----RQDWQLSLRR 375
QY  427  QHYSFFGVE---QANWDNFTLEGGVRVBKQKASIRYD---KALIDRENYKQ-ELPDIGA 479
Db  376  DDNQFGVHDTGSAAW-GYALSADALRF-----TVNYGTAFKAPTFLNELYYPDYGNDPLDA 429
QY  480  HROQTARSPALSGNNWYFTTQHKLSLTASHQERLPSQELYAHG-KHVAINTFE-----VG 532
Db  430  EYRSLEVLGSG-----THGWHGWA VNAFRTNVDDLIG 462
QY  533  N-----KHLNKRSNNIETALGVGEDRWQYNL-ALY---RNRFGNYIYAQTL 575
Db  463  NDPRPAPGRPWGPNNAIDEARIGVELVLGSGWLGDWNNANATFLDPQNRSGG-----V 516
QY  576  NDG-----RGPKSIEDDSRMKLVRYNQSGADPYGAGEGRIYFKPTPRYRIG-----VSGD 624
Db  517  NDGNEULPRARRMNFLEDRRPERLS-IGASVH-AEGRRYDDPANKVLGAYATLDURSE 574
QY  625  Y-----VRGLKNLPSLPGREDAYG-NRP 647
Db  575  YRLNDEWRLOGRITANLFG-ADYETAYGYNOP 604

```

```

RESULT 12
US-09-328-352-4834
; Sequence 4834, Application US/09328352
; Patent NO. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4834
; LENGTH: 707
; TYPE: PRP
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4834

```

Query Match	4.8%;	Score 193;	DB 4;	Length 707;
Best Local Similarity	19.0%;	Pred. No. 1.8e-09;		
Matches 162;	Conservative 126;	Mismatches 312;	Indels 252;	Gaps 35;
Qy	6	LKPIVLSLLINTPLAQAHETEQSGVLETVVVKSRPRATSGLLHTSTASDKTIISGDT	65	
Db	9	LQPLWVAMIAVHSHGLVPFAKSEKNDADTNLHSLAPIVVTAAQQ-----NDANGLIIVHADP	64	
Qy	66	LRQ-----KANVLGBALDGVPGIHAISQYGGASAPVIRGQTRRIKVLNHHGET-GDMADF	120	
Db	65	KQIQPVPAFDGADVLQSGIMGFNSIQSGTNGDVTFRGMFGSRIKILTDGTENLGACPNR	124	
Qy	121	SPDHAIWDTALTSQQVEITLRGPVTLVLYSSGNAGVLDVADGKIPEKMPENGVSGBELGRL	180	
Db	125	MDAFTSYLSPSRYDRISVIKGPQIVQANTGSAATV-----LFRQP-----	166	
Qy	181	SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAVPRYRNLRKLPSRPRFANGOHRA	240	
Db	167	-----EKITS-----EKPYRQASVLLGSGRI-----DHNVEA	195	
Qy	241	VLGWRNRRYRRTYSDRRDQYGLPAHSHYDDCHADII--WOKSLINKRYLQLYPELLUTE	297	
Db	196	AVGDEKK-YIRLNAIR-----SESNSYQDGDGNTVPSAWKKW-----	231	
Qy	298	EDVDYDNFGLSCGFHDDDDAHAAHANGKPWIDLNRKYE-----	336	
Db	232	-----NVDVALGFTPDENT-----WVEITGCKSDGESLYAGRSMDSQFARESILG	276	
Qy	337	LRAEWKQPPGPFREAL--RVHLNRNDYHHDEKA-----GDVAENFFNN	376	
Db	277	LRFEKKNITDVIKKITEGOWNYSYNDHIMDNFSLRTPPLVEMNHGGMTLMPNAMAMOVTR	336	


```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: E. coli CirA protein amino acid sequence
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia Coli
US-09-098-082-7

Query Match 4.78; Score 188.5; DB 3; Length 663;
Best Local Similarity 20.5%; Pred. No. 4.4e-09; Indels 301; Gaps 44;
Matches 176; Conservative 110; Mismatches 270;

Qy 6 LKPIV-----LSIIILNTPLLAQAHETEVSGLTETVVVKGSRPRATSGLLHTSTASDKI 60
Db 4 LNPFRVGLCLSAISCAWPLAVDDDG-----ETWVTASS-----VEQNKLDPASISV 53
Qy 61 ISGDTLQKAV-NLGDALDGVPGHSAQYGGASAPVIRG-----QTGRRIK----- 106
Db 54 ITQDLQKPVQNLKQVLEKEVPGVQLTNEGDRKGVSGIRGLDSYTLILVDGKRVNSRNA 113
Qy 107 VLNHGEGTGMADSPDHAIMVDTALSGQVEILLRGPVTLTYSSGNVAGLVADGKIPEK 166
Db 114 VFRN-----DFDLNW-IPVDSI--ERIEVVGPMSSLYGSDALGGVNIITKIGQK 163
Qy 167 MPENGVSGLGL-----RLSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAV 217
Db 164 W-----SGTVTDVTIQHRRDRGPTYNGQFFTSGLIDGV-----LGM 201
Qy 218 PRYENL-KRLPDSPRFA--NGQRAVLGWRKFRYRTYSRRDOYV-LPAHSHEY--- 269
Db 202 KAYGLAKREKDDQNSTTDTGTPTRIEGFSR-----DGNVEFAWTPNQNHDTAG 254
Qy 270 -----DCHADIIWQKSLINKRY-----LQYPHLLTBEDVDYDNPGLSCG 310
Db 255 YGDFQRQDSLSLKNRLERQVSVSHNGRWDTYSELKTY-----GEKVENKNPGNSSP 309
Qy 311 FHDDDDAHAAHAKPMIDLR--NKRVELRAEWKQPPGFEALRVHLNRNDYHHDKAGD 368
Db 310 I-----TSESNTVDGKYTLPLAINQFLTGVGEW-----HD-KLSD 345
Qy 369 AVENPFNQNTQARIELRHQPIGRIGKSGWQYLGQKSSALSATSEAVKQFMLLDNKVQH 428
Db 346 AV-----NLT-----GGTSSKTSAS-----QY 362
Qy 429 YSPFGVEQVNDNFTLEGVREVEKASIRYDKALIDRENYKQPLDGLGAHRTARSFA 488
Db 363 ALFVEDEWRIEPLALTTGVKMD-----DHETYGHSWSPRAYLVYNADTVT 409
Qy 489 LSGNWFYTPQHKLSLTASHQRLPSTQELYAHGKHVATNTPF-----VGNKHLNKNRSNN 543
Db 410 VKGW-----ATAF--KAPSLQL---SPDWTSSNCRGACKIVGSPDLKPETSES 454
Qy 544 IELALGYEGDR-W-----QYNLALYN-----RFGNYIYAQTLNDGRG 580
Db 455 WELGLYNGEGLGVESSVTVFRNDVKDRISRTSDVNNAAFGYQNFVGFETGANGRR 514
Qy 581 PKSIDDDSEMKLVRNQSGADFYGAEGBIYFKPTPRYRIGVSGDYVRGRLKNLPSLPGRE 640
Db 515 IPVFS-----YVNVKARNQGVETELKTFPNDENKLSINYTYNDR----- 555
Qy 641 DAYGNRPFIADDDQNAVPARLGFHLKASLTDRIDANLDYY---RVFAQNKIARVETR 697
Db 556 -----DVSNGENKPLSLDLPFHL--ALED-----WSFYVSGHYTGOKRADSATAK 597
Qy 698 TPGHHML-NLGA-----NYYRNRTRYGEMNVTYKADNLLNQSVVAHSSFLS 741
Db 598 TPGGTYIWTGAWQVTKDKVLKAGVLNLGDKTANGTLDW--KPD--LSRDDISYN----- 649
Qy 742 DTPQMGSRFTGGVNVKF 758
Db 650 ---EDGRRYFMAVDYRF 663

```

```

RESULT 16
PCT-US95-06994-7
; Sequence 7, Application PC/TUS9506994
; GENERAL INFORMATION:
; APPLICANT: Children's Hospital & Medical Center
; APPLICANT: University of Washington
; APPLICANT: Washington State University Research Foundation
; APPLICANT: TARR, PHILLIP I
; APPLICANT: BILGE, SIMA S
; APPLICANT: BESSER, THOMAS E
; APPLICANT: VARY JR, JAMES C
; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
; STREET: SUITE 2800, 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: WA 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: 07-JUN-95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BRODERICK, THOMAS F
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: CHOR-18591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682 8100
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrpA amino acid sequence,
; DESCRIPTION: wherein Xaa residues represent gaps
; DESCRIPTION: introduced to facilitate best alignment with
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: VIBRIO CHOLERAE
; PCT-US95-06994-7

Query Match 4.5%; Score 182; DB 5; Length 718;
Best Local Similarity 21.1%; Pred. No. 2.1e-08;
Matches 160; Conservative 85; Mismatches 292; Indels 222; Gaps 34;

Qy 1 MAQTTLKPIVLSI---LLINTPLLAQAHETEVSGLTETVVVKGSRPRATSGLLHTSTAS 57
Db 1 MSRFNPSVLSLTVTLGLMFSASXAPADATKTDFTVMVTAAGYQA-----VIQNPAS 54
Qy 58 DKTIISGDTLRQKAV-NLGDALDGVPGHSAQY-GGASAPV-IRG-----Q 100
Db 55 ISVISREDLESRYRVDVTDALASVFCVTKYXGTGGDTTDSIRGWSNYTLILVDGKRQ 114
Qy 101 TGRRIKVLNHGEGTGMADSPDHAIMVDTALSGQVEILLRGPVTLTYSSGNVAGLVADVAD 160
Db 115 TSGRTRP-NSDGFIEQGWLPPLQAI-----ERIEVIRGPMSTLYGSDAIGGVNIIT 166
Qy 161 GKIPKWPENGVSGELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAVRY 220
Db 167 RKDQQW-----SGNVOLSTVQEN-----RAGDQSQSANF 197

```


QY 577 ---DGRGPKSIEDDSEMKLVRYN-QSGADFYVARGSEIYF-----611
 Db 715 KNGTGKANYGHNAQNAKLVUNTAQDLDFNGLWKRIPYGYWATFAYNRVKKVQDKINAG 774
 QY 612 ---KPTPRYRIGVSGD-----YVRGLKNLPSLPQREDAYGNRPF 648
 Db 775 LASVSSYLDFAIQPS-RYIIGLYDHPSTNGINTMTQSKAKSQNELLGKR-ALGNN-- 830
 QY 649 IAQDDQNAFVPAARLGFHLKASLTDRIDANLDYRVFAQNKLYETRTFGHHMLNLA 708
 Db 831 -SRDVKSTRKLTTRA---WHI-----LDVSGYTMANK-----NIMRLGI 865
 QY 709 NYRNRTRYGEMWYVK-ADNLLN--QSVYHSSFLSDTPQMGSRFTGGVNVKE 758
 Db 866 YNLFNRYVTWEAVRQTAQAVNQHVNGSVTRYAAS-----GRNYTLTLEMKF 914

RESULT 18
 US-08-478-435-11
 ; Sequence 11, Application US/08478435
 ; Patent No. 5922323
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/478,435
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/337,483
 ; FILING DATE: 08-NOV-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/175,116
 ; FILING DATE: 29-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/148,968
 ; FILING DATE: 08-NOV-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 914 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-478-435-11

Query Match 4.5%; Score 180.5; DB 2; Length 914;
 Best Local Similarity 19.7%; Pred. No. 4.5e-08;
 Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;
 QY 5 TLKPI-----VLSILLINTPLLAQH---ETEQSVG-----LETVTWVG-KSRP 44
 Db 2 TKKPYPRLSIIISCLLISCVVKAETQIKOTKEAISSEVDTQSTEDSELETISVTAEKIR 61
 QY 45 RATSGLLHTSTADKII--SGDTLRQKAVNLGDALDVGPIHASYQGGASAPV-IRGT 101
 Db 62 RQNEV---TGLGKIIKTSBSISREQVNLHDLTRYDFGISVVEQGRGASGYSIRGMD 117
 QY 102 GRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPVTLT---YSSGNVAGLDVA 159
 Db 118 RNRV-----ALLVDGLPQTQSVVVCSPINVARSGYSGTGAINIEIYE 158
 QY 160 DGKIPKMPENGVSGLGLRLSSGNLEKLTSGGINIGL-KNFVLHTEGLYR-----210
 Db 159 NVKAVE-ISKGGSSSEYNGALAGSVTFQSKAADILEGKDSWGIQTKNAVSSKNKGFTH 217
 QY 211 -----KSGDY---AVPRYRNLKRLPDSPPRFANGQHR-AVLGWRKRFYRRYTSRRDOY 260
 Db 218 SLAVAGKQSGPDGVALYTORN-----SIETQVHKDALKGVS--YHRLIAPEDQ- 265
 QY 261 GLPAHSHEYDDCHADIIWOKSL-INKXYLQLYPHLLTEEDVY-----DNP-----305
 Db 266 --SAYFVMQDECPKDDYNSCLPFAKRPAILSSQRETSVSDYTGANRIKPNPMKYESQS 323
 QY 306 -GLSCGFHDDDDAHAAHNGKWIIDLNRKRYELRAEWKQPPGFEALRVHLNRNDYHDE 364
 Db 324 WFLRGGHFSEQ---HYIGG--IFFTQOKFDIR--DMTFPAY-----LRSTE 364
 QY 365 KAGDAVENFNNQTQNA--RIELRHQPIGR-----LKGSGVQVYLGOKSSA 408
 Db 365 KRDDSSGSFVPKQDYGAYQRIE---DGRGVNYASGLYFDBHHRKQRVGIEVIYENKNK 419
 QY 409 LSATSEAV---KOPMLDNKVOH-----YSPFGVE-----435
 Db 420 AGIIDKAVLSANQNIILDSYNQHTCSLYPNPSKNCRPTRDKPYSYSHSDNRVYKEKN 479
 QY 436 -----QANW-----DNFT-----LEGVVRVEKQKASIRYDKALIDR 466
 Db 480 MLQNLLEKKIQNWLTQIVFNLFDDFTSALQKRDYLTFRVTATAKSI---EKANETR 536
 QY 467 ENYKQ-----PLPDIG-----AHRQTARSAFSGNWNYSFPPQKLSL-----503
 Db 537 RNYKKQPYLYPKPTVGVVQDHCYKNSNRYRDCVKRLIKGKNYFFAARNMALGKTV 596
 QY 504 -----TASHOE-----RSTQEL 517
 Db 597 DLGLGIRYDVSRTKANESIISVGPKNPSWNTGIVIKPTEWLDLSYRLSTGFNPSFAEM 656
 QY 518 YAHGKHVATNTTFVGNKHLNKNRSNNIELALGYEGDRWOYNLALYNNRFGNYI-YAQTJN 576
 Db 657 Y--GWRYGNNSEVYVGVKPKPETSRLNQBFGLKALGDFGNIEISHFSNAYRNLIAPAE 714
 QY 577 ---DGRGPKSIEDDSEMKLVRYN-QSGADFYVARGSEIYF-----611
 Db 715 KNGTGKANYGHNAQNAKLVUNTAQDLDFNGLWKRIPYGYWATFAYNRVKKVQDKINAG 774
 QY 612 -----KPTPRYRIGVSGD-----YVRGLKNLPSLPGRBDAYGNRPF 648
 Db 775 LASVSSYLDFAIQPS-RYIIGLYDHPSTNGINTMTQSKAKSQNELLGKR-ALGNN-- 830
 QY 649 IAQDDQNAFVPAARLGFHLKASLTDRIDANLDYRVFAQNKLYETRTFGHHMLNLA 708
 Db 831 -SRDVKSTRKLTTRA---WHI-----LDVSGYTMANK-----NIMRLGI 865
 QY 709 NYRNRTRYGEMWYVK-ADNLLN--QSVYHSSFLSDTPQMGSRFTGGVNVKE 758
 Db 866 YNLFNRYVTWEAVRQTAQAVNQHVNGSVTRYAAS-----GRNYTLTLEMKF 914

```

RESULT 19
US-08-337-483-11
; Sequence 11, Application US/08337483
; Patent No. 5922562
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESS: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-337-483-11

Query Match 4.5%; Score 180.5; DB 2; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI---VLSILLINTPLLAQAH---ETEQSVG-----LETVTVVG-KSRP 44
Db 2 TKRPFRLISICLLISCVKAEQTQSKDTKEALISSEVDQTQSTSELETISVTAEKIRD 61
Qy 45 RATSGLLHTTASDKII--SGDTLRQKAVNLGDALDVGPGIHASQYGGGASAPV-IRGQT 101
Db 62 RKDNEV---TGLGKIITSEISREQVLNIRDLTRYDGISSVVEQGRGASSGYSIRGMD 117
Qy 102 GRRIKVLNHCETGDMADFPDHAIMVDTALSSQVEILRGPTLL--YSSGNVAGLVDA 159
Db 118 RNRV-----ALLVDGLPQTQSVVQSPVARGSGYGTGAINETYE 158
Qy 160 DGKIPEKMPENGVSSELGLRLSSGNLEKLTSGGINIGLG-KNFVLHTEGLYR----- 210
Db 159 NVKAVE-ISKGGSSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTKNAYSSKNKGFTH 217
Qy 211 -----KSGDY---AVPRYNLKELPDPRFRANGQHR-AVLGWRKRFYERTYSDRDQY 260
Db 218 SLAVAGKGGGEGVAYTYQRN-----SIETQVHKDALKVQS--YHRLAKPEDQ- 265
Qy 261 GLPAHSHRYDDCHADIIWQKSL-INKRYLOLYPHLLTEEDVDY-----DNP----- 305
Db 266 --SAFVNVQDCPKPDDYNSCLPFAKRAPIILSSQRETYSVDYTGANRIKPNPMKYESQS 323

```

```

Qy 306 -GLSCGFHDDDDAHAAHNGKFWIDLRNKRKYLAEAWKQPPGFEALRVHLNRDYYHDE 364
Db 324 WFLRGYHFSEQ---HYIGG--IFEFTQOKFDIR---DMTFPAY-----LRSTE 364
Qy 365 KAGDAVENFENNOTQNA--RIELRHQPIGR-----LKGSGWQYQYLQKSSA 408
Db 365 KRDDSSGSFYPKQDYGAYQRIE---DGRGVNVSAGLYFDEHHRKQKRVGLIYVENKKN 419
Qy 409 LSATSEAV---KQMLLDNKVQH-----YSFGVE----- 435
Db 420 AGIIDKAVLSANQQNIILDSYMOHTHCSLYPNPKNCRPTRDKPYSYSHSRNVYKEHN 479
Qy 436 -----QANW-----DNFT-----LEGVVRVEKQKASIRYDKALIDR 466
Db 480 MLQNLLEKLIQONWLTHQIVFNLGFDFTSALQHKDYLTRVTATAKIS---EKANETR 536
Qy 467 ENYKQ-----PLPDIG-----AHRQTARSPALSGNWIYTPQHKLSL----- 503
Db 537 RNYGKKQPYLYPKPTVGVVQDHCDYKGNSSNYRDKVRLINGKNYFFAARNMALGKYV 596
Qy 504 -----TASHOE-----RFPSTOBL 517
Db 597 DLGLGTRYDVSRITKANESTISVGKFNFSWNTGIVIKPTWLDLSYRLSTGPRNPSFAEM 656
Qy 518 YAHGKHVATNTFEVGNKHLNKNRSNNIELALGYEDRWQYNNALYRNRFNGVI-YAQTIN 576
Db 657 Y--GWRYGNNSEVYVGVKPKETSRNQEFGLAKGDFGNIETSHFSNAYRNLIAPAEELN 714
Qy 577 ---DGRGPKSIEDSEMKLVRYN-OSGADFYGAEGEIVF----- 611
Db 715 KNGTGKANYGYHNAQNAKLGVGNITRAQDFNGLNGKKEIPYGYATATAYNREVKVKQKINAG 774
Qy 612 -----KPTPRYRIGVSGD-----YVGRKLNPLSPDGRDAYGNRPF 648
Db 775 LASVSSYLEDAIQPS-RYIIGLYDHPSTNWTGINTMTQSKAKSONELLGKR-ALGNN-- 830
Qy 649 IAQDDQNAVPAPARLGFHLKASLTDRIDANDLYRVEFAQNKILARVETTPGHMLNLGA 708
Db 831 -SRDVKSTRKLTRA---WHI-----LDVSGYINWAK-----NIMRLGI 865
Qy 709 NYRNRTRYGEWNWYVK-ADNLLN--QSVYAHSSFLSDTFQMGSRSTGGVNVKVF 758
Db 866 YNLFNYEYVTVWEAVRQTQAGAVNQHVGVSYTRYAAS---GRNYTLTLEMKF 914

RESULT 20
US-08-478-373-11
; Sequence 11, Application US/08478373
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESS: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25

```



```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-373-11

Query Match      4.5%; Score 180.5; DB 2; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

QY 5 TLKPI-----VLSILINTPLLAQH---ETEQSVG-----LETVTVG-KSRP 44
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2 TKKPYRLSIICLLISCVKAEIOSIKDTKEAISSEVDSTOSTEDSELETISVTAEKIRD 61
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 45 RATSGLLHTSTASDKII--SGDTLQKAVNLGDALDGVPGIHASQYGGASAPV-IRGQT 101
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 RQDNEV---TGLGKIITSEISREQVNLINRDLTRYDPGISVVEQGGAGSSGYSIRGMD 117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 102 GRRIRKVLNHHGTGDMADSPDHAIMVDITALSQQVEILRGPVTLI--YSSGNVAGLVDA 159
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 RNRV-----ALLVDGLPQTQSVYVQSPVAVRSYSGTGGAINEIYE 158
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 DKGIPEKMPENGVSGLGLRLSSGNLEKLTSGINIGLG-KNFVLHTEGLYR----- 210
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 NVKAVE-ISKGSSEYNGALAGSVTFQSKAADILEGDKSWGIGTKNAYSSKNKNGFTH 217
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 211 -----KSGDY---AVPRYRNLKRLPDSRRFANGQHR-AVLGWRKRFPYRRTYSRRDQY 260
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 SLAVAGKQSGFDGVALIYTORN-----SIETQVHKDALAGVOS--YHRLIAKPEQ- 265
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 261 GLPAAHSHVEDDCHADIIWQKSL-INRKYQLPHLLTEEDVY-----DNP----- 305
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 --SAYFMQDECPKPDYNSCLFPKRAIPAILSSQRETVSVSDYTGANRIKPNPMKYESQS 323
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 -GLSCGFHDDDAHAHAHKGKWDILNRKRYELRAEWKQPPFGFEALRVHLNRNDYHDE 364
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 WFLRGYHHSQ---HYIGG--IFBETQKFDIR---DMTFAY-----LRSTE 364
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 365 KAGDAVENFNNQTONA--RIELRHQPIQR-----LKGSGVGYLQKQSSA 408
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 365 KRDDSSGSFPKQDYGAYORIE-----DGRGVNYSGLYFDBHHRKQRVGIEYVENKNK 419
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 409 LSATSEAV---KQPMLLDNKVQH-----YSFFGVE----- 435
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 420 AGIIRKAVLSANQNIILDSYMOHTHCSLYPNPSKNCRTDRKPYSYHSDRNVYKEKH 479
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 436 -----QANW-----LEGGVYVEKQKASIRYDKALIDR 466
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 480 MLQINLEKKIQNWLTHTQVFNLFDPDDFTSALQHKDYLTRVTATAKSIS---EKANETR 536
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 21

```

US-08-474-671-11
; Sequence 11, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mardin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973

```

```

QY 467 ENYIKO-----PLPDLG-----AHROTARSGFALSQNMWYFTPQHKLSL----- 503
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 537 RNYGKKQPYLYPKFTVGVVQDHCDYKGNSSNVYRCKVRLIKGKNYFAARNMALGKTV 596
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 504 -----TASHQE-----RLPSTQEL 517
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 597 DLGLGIRYDVSRTKANESTISVGKFKNFWSNTGIVIKPTWMLDLSYRLSTGFRNPSFAEM 656
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 518 YAHGKHVATNTFEVGNKHLNKRSSNNIELALGVEGDRWQVYNLALYENRFQNYI-YAQTIN 576
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 657 Y--GWRYGNNSEVYGVKFKPETSRAQEFCLALKGDFGNIEISHFNSAYRNLIAPAEELN 714
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 577 ---DGRPKSIEDSEMULVRYN-QSGADFYAGABSIYF----- 611
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 715 KNGTGKANYGYHNAQNAKLGVNITAQDLDFNGLWKRIPIYGVWATFAYNRVKVKDQKINAG 774
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 612 -----KPTPRYRIGVSGD-----YVRGLKNLPSPLPREDAYGNRPF 648
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 775 LASVSSYLFDAIQPS-RYIIGLYGHPSTWNTGINTFTQSKAKSQNELLGKR-ALGNN-- 830
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 649 IAQDDQNAKPRVPAARLGFHLKASLTDRIIDANLDYRVYFAQNKLYRYETRTPGHMLNLGA 708
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 831 -SRDVKSTRKLTRE--WHI-----LDVSGYVMANK-----NIMLRIGI 865
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 709 NYRNTRYGEWNYVK-ADNLLN--QSVYAHSSFLSDTPQMGRSFTGGVNVKF 758
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 866 YNLFNRYVTWEAVROTAQGANVQHQNVGSYTRYAAS-----GRNYTLTLEMKF 914
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:Vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-474-671-11

Query Match          4.5%; Score 180.5; DB 3; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI-----VLSILLINTPLLAQAH---ETQSVG-----LETVTVVG-KSRP 44
Db 2 TKKPVFRLSIISCLLISCVKRAETQSIKDTKEAISSEVDQTQSTEDSELETISVTAEKIRD 61
Qy 45 RATSGLLHTSADKII--SGDTLRQKAVNLGDALDGVPGIHASQYGGASAPV-IRGQT 101
Db 62 RKDNREV-----TCLGKIIITSISREQVLNIRDTRYDPGISVVEVQGRGASSYSIRGMD 117
Qy 102 GRRIKVLNHHGETGMDPSPDHAIMVDTALSQQVEILRGPVTL--YSSGNVAGLVDA 159
Db 118 RNRV-----ALLVDGLPQTQSYVVQSLVARSYGSGTGAINIEIYE 158
Qy 160 DGKTIKPKMPENGVSGLRLSSGNLEKLTSGINIGLG-KNFVLHTEGLYR----- 210
Db 159 NVKAVE-ISKGSSESSEYNGALAGSVTFQSKSAADILEGDSKSGTQTKNAYSSKNKGFTH 217
Qy 211 -----KSGDY---AVPRYRLKRLPDSRPFANGQHR-AVLGWRKRYRRTYSRRRQY 260
Db 218 SLAVAGKGGDGVDAIYTORN-----SIETQVHKDALKGVQS--YHRLIAKPEDQ- 265
Qy 261 GLPAHSHEYDDCHADIWQKSL-INKRYLQLYPHLLTEEDVDY-----DNP----- 305
Db 266 --SAYFVQDECPRPDDNSCLPAKRPAILSSQRETVSVSDYGANRIKPNPKYESQS 323
Qy 306 -GLSCGFHDDDAHAHANGKPIDLRNKRVELRAEMKQPPFGEALRVHLNRNDYHDE 364
Db 324 WFLRGGVHFSQ---HYIGG-IFEFTQCKDIR--DMTFPAY-----LRSTE 364
Qy 365 KAGIAVENFFNQTQNA--RIELHQPIGR-----LKSGWGVQYLGQKSSA 408
Db 365 KRDDSSGSFYPKQDYGAYQRIE-----DGRGVNVSAGLYFDEHHRKQGVGIEIYENKNK 419
Qy 409 LSATSEAV-----KQPMLLDNKVQH-----YSPFGVE----- 435
Db 420 AGIIDKAVLSANQQNIILDSYMQHTHCSLYPNPSKNCRPTDKPYSYVHSRNVYKEKH 479
Qy 436 -----QANW-----DNPT-----LEGGVRVEKQKASIRYDKALIDR 466
Db 480 MLQNLLEKKIQONLTHQIVENLGFDDFTSALOHDVLTTRVTATAKSIS---EKANETR 536
Qy 467 ENYKQKQ-----PLDGLG-----AHRQTARSFALSGNWWYFTPOHKLSL----- 503
Db 537 RNYGKQPYLPKTVGVFVVQVQDHCYKGNSSNYRDKVRLIKGNKYVFAARNMALGKYV 596
Qy 504 -----TASHQE-----RLPSTQEL 517
Db 597 DLGLGIRYDVSRTKANESTISVGKFKNFSWMTGIVIKPTEWLDLSYRLSTGFRNPSFAEM 656
Qy 518 YAHGKHVATNTPFVGNKHLNKRNNIELALGYGDRQWYNLALRYRNFNGYI-YAQTLN 576
Db 657 Y--GWRVYGGNNSEYVVGKFKPETERNQEFGLAKGDFGNIIEISHFSNAYRLNIAFAEELN 714
Qy 577 ---DGRGPKSTEDDSEMKLVRYN--QSGADFYVAGSELYF----- 611
Db 715 XNGTGKANYGHNAQNAKLVGVNNTAQLDNFGLWKRLPYGHWATFAYNRVKKVQDKINAG 774
Qy 612 -----KPTPRYIGVSGD-----YVRGLKLNPLSILPREDAYGNRPF 648

; 775 LASVSSYLFDALQPS-RYIIGLGYDHPNSWTGINTMTFQSKAKSQNELLGKR-ALGNN-- 830
; 649 IAQDDONAPRVPAARLGHPLKASLTDRIDANLDYRVFAQNKLYRYETRTPGHHMLNCA 708
; 831 -SRDVKSTRKLTTRA---WHI-----LDVSGYTMANK-----NIMLRIGI 865
; 709 NVNRNTRYGEWNWYVK-ADNLLN--QSVYAHSSFLSDTPQMGSRFTGQVNVVKF 758
; 866 YNLFNRYVTWEAVRQTAQGVNQHVGVSYTRIAAS-----GRNYTLTLEMKF 914

RESULT 22
US-08-483-577A-11
; Sequence 11, Application US/08483577A
; Patent No. 6015688
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08483,577A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 914 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-483-577A-11

Query Match          4.5%; Score 180.5; DB 3; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI-----VLSILLINTPLLAQAH---ETQSVG-----LETVTVVG-KSRP 44
```


Db 537 RNYKKQPYLPKPTGVFVVDHCDYKGNSSNYEDCKVRLIKGKYYFAARNMNLGKYY 596
Qy 504 -----TASHQE-----RIPSTQEL 517
Db 597 DLGLGIRYDVSRTKANESTTISVGKFKNFSWMTGIVIKPTWELDLSYRLSTGFRNPSFAEM 656
Qy 518 YAHGKHVATNTFEVGNKHLNKRNNIELALGYEGDRWQYNLALYRNRFGNYI-YAQTIN 576
Db 657 Y--GWRYGNNSEVYVVKFPETSRNQEFGLAKGDFGNIHSHFSNAYRNLTAFABELN 714
Qy 577 ---DGRGPKSIEDDSEMKLVRYN-QSGADFYGAGEIYF----- 611
Db 715 KNGTGKANYGYHNAQNAKLVGNITAOIDENGLWKRIPYGYWYATFAYNRVVKVDQKINAG 774
Qy 612 -----KTPRYRIGVSGD-----YVRGLKNLPSLPGREDAYGNRP 648
Db 775 LASVSSYLFDAIQFS-RVILGVDHPSNTWGMTQSKAKSQNELKGR-ALGN-- 830
Qy 649 IAQDDONAPRYPAARLGFHLKASLTDRIANLDYRVFAQNKLYARTPTPGHMLNLA 708
Db 831 -SRDVKSTRKTRA---WHI-----LDVSGYTMANK-----NIMLRGI 865
Qy 709 NYRNTRYGWNVYK-ADNLLN--QSVYAHSSFLSDTPQWGRSFTGCVNVKF 758
Db 866 YNLFNRYVTWEAVRQTAQAGAVNOHNVGSYTRYAAS-----GRNYTLTLEMKF 914

RESULT 25
US-08-649-518-11
; Sequence 11, Application US/08649518
; Patent No. 6361779
; GENERAL INFORMATION:
; APPLICANT: Loomsore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,518
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 914 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-649-518-11

Query Match 4.5%; Score 180.5; DB 4; Length 914;
Best Local Similarity 19.7%; Pred. No. 4.5e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

Qy 5 TLKPI-----VLSILLINTPELLAAH---ETEOSVG-----LETVTVVG-KSRP 44
Db 2 TKKPYFRLSIIICLLISCYVKAETOSIKDKAISSEVDTQSTEDSELETISVTAEKIRD 61
Qy 45 RATSGLLHTSTASDKII--SGDTLRQKAVNLGDALDGVGIIHASQYGGGASAPV-IRGQT 101
Db 62 RKNV-----TGLGKIITSEISREQVNLIRDLTRYDFGISWVEQGRGASSGYSIRGMD 117
Qy 102 GRIKIVLNHGETGMADFPDPAHIMWDTALSOQVBEILRGPTVLL--YSSGNVAGLVDA 159
Db 118 RNRV-----ALLVDGLPQTQSYVQSPVARSYSGTGTAINIEYE 158
Qy 160 DGKIPEKMPENGVSGLRLSSNGLEKLTGGINIGLG-KNFVLHTEGLYR----- 210
Db 159 NVKAVE-ISKSSSYNGALAGSVTFQSKAAADILEGDKSWGIGTKNAYSNNKNGFTH 217
Qy 211 -----KSGDY---AVPRYENLKLPDPRFANGQHR-AVLGWRKRFYRTYSRRDQY 260
Db 218 SLAVAGKQGGFGDGAIVTQRN-----SIETQVHKDALXGVS--YHRLIAKPEDQ- 265
Qy 261 GLPAHSHEYDDCHADIIWQKSL--INKRYLQYPLHLLTEEDVY-----DNP----- 305
Db 266 --SAYFMQDECPKDDYNSCLPFAKRPAILSSQRETYSVDYTGANRTPKPMKYESQS 323
Qy 306 -GLSCGFHDDDAHAHAKPWIDLRNKRYELRAEWKQFPFGFEALRVHLNRNDYHDE 364
Db 324 WFLRGVHFSEQ---HYIGG--IFEETQCKFDIR---DMTFPAY-----LRSTE 364
Qy 365 KAGDAVENFNNOTQNA--RIELRHOPGR-----LKGWGVQVLGQKSSA 408
Db 365 KRDDSSGSFYPKQDYGAYQRIE-----DGRGVNYSGLYFDEHHRKQRYGIEIYENKN 419
Qy 409 LSATSEAV---KOPMLLDNKKVQH-----YSFGFVE----- 435
Db 420 AGIIDKAVLSANQONILDSYMOHTHCSLYPNPSKNCRPTRDKPYSYTHSDRNVYKEKN 479
Qy 436 -----QANW-----DNFT-----LEGGVVEKQKASIRYDKALIDR 466
Db 480 MLQINLEKTIQQNWLTHQIVNGLGDFDFTSALQHKDYLTTRVTATAKSIS---EKANETR 536
Qy 467 ENYKQ-----PLPDLG-----AHRQTARSFALSNGNMYFTPHKLSI--- 503
Db 537 RNYKKQPYLPKPTGVFVVDHCDYKGNSSNYRDCVKRLIKGKYYFAARNMNLGKYY 596
Qy 504 -----TASHQE-----RIPSTQEL 517
Db 597 DLGLGIRYDVSRTKANESTTISVGKFKNFSWMTGIVIKPTWELDLSYRLSTGFRNPSFAEM 656
Qy 518 YAHGKHVATNTFEVGNKHLNKRNNIELALGYEGDRWQYNLALYRNRFGNYI-YAQTIN 576
Db 657 Y--GWRYGNNSEVYVVKFPETSRNQEFGLAKGDFGNIHSHFSNAYRNLTAFABELN 714
Qy 577 ---DGRGPKSIEDDSEMKLVRYN-QSGADFYGAGEIYF----- 611
Db 715 KNGTGKANYGYHNAQNAKLVGNITAOIDENGLWKRIPYGYWYATFAYNRVVKVDQKINAG 774

Qy 612 -----KPTPRYRIGVSGD-----YRGLKMLPSLPGREDAYGNPEP 648
 Db 775 LASVSYLFDIAQPS-RYIIIGLYDHPENTWINTMTQSAKQSNELLGR-ALGNN-- 830
 Qy 649 IAQDDQNAAPRVEAPARGLKASLTDRIDANLDYRVRFAQMKLARYETRTPGHMLNLA 708
 Db 831 -SRQVSKSTRKLTTRA---WHI-----LDVSGYYMANK-----NIMLRGI 865
 Qy 709 NYRNRTRYGEWNNYVK-ADNLLN--QSVYAHSSFLSDTPQGRSFTGGVNVKF 758
 Db 866 YNLFNRYVTWEAVRQTAQAGVNOHQNVGSYTRYAAS-----GRNYTLTLEMKF 914

RESULT 26
 US-09-252-991A-30809
 ; Sequence 30809, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 30809
 ; LENGTH: 756
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-30809

Query Match 4.4%; Score 179; DB 4; Length 756;
 Best Local Similarity 20.7%; Pred. No. 4.5e-08;
 Matches 168; Conservative 104; Mismatches 266; Indels 272; Gaps 42;
 Qy 8 PIVLSILL---INTPLLAQAHTESVGLTETVVVGKSRPRATSGLLHTSTASDKIISGD 64
 Db 25 PLALSLSLFAFSAPALA-ADPEVQ-----MVVIGSRAPTRISELPGTVVWIER-EQLD 77
 Qy 65 TLROKAVNLGDALDG-VPGIHASQVGGASAPVIRGQTRRIKVL---NHGEGTDMAD 119
 Db 78 QQTQAGVPLKALGQLIPGLDICSQGRTNQNL---GRSLVMDGVLSNRSRISQ 134
 Qy 120 FSPDHAIWDTALSOQVEILRGVETLLYSNGVAGLVADVADGKIPEKMPGVSG----- 174
 Db 135 FD-----SIDPFNIEVWSG-ASAVYGGGATGGIINIVTK-----GVGDTFRN 180
 Qy 175 -ELGLR-----LSSGNLEKLTSGGINIGLGNFVLHTEGLYRKSGDYAV 217
 Db 181 TELGASRGQSHEHDHLRAAQISGN--DLFNGRLAIYQK-----GAAYDGSQDVL 233
 Qy 218 P-----RY-RNLKRLPSPRRFANGQRAVLGWRKRFYRRTYDRDRDQYGLPAHSHEY 269
 Db 234 TDITQTDQVNRSDVLMGSLGTFANG-HSLDLG----- 266
 Qy 270 DDCHADIWQKSLINKRYLYQLPHLLTEEDVDYD-NPGLSCGFHDDDDAHAHNKPWI 328
 Db 267 -----LQY-----DSGVDGDRGLDGRNFDALGRAPYSIKGV 301
 Qy 329 DL-----RNKRYELRAEWKOPFPQGEALRYHLNRNDYHDEKAGDAVENFPNNTQONARIE 384
 Db 302 DLREPESKEHQFNATHPA---EVLGHDLYLQAYYRNEK-----MAFNPFPT 346
 Qy 385 LRHQPTRLKSGVQVLOKSSALSATSEAVKQPMLLNKVQHSFFVEQA---NWDN 441
 Db 347 IRYNTGAI--NYGTSY-----YSASQ-----QDTDYGMKALVKTWER 384
 Qy 442 FTLEGGVRVBEKQKAS-----IRYDKALIDRENYVQKPLDLAGHQTA 484

Db 385 ASLTGVLDLREKFTSDQMLFNLPLAASGGLVASEQAKLGR-----YPDIDT---DS 434
 Qy 485 RSFALSGNWFYTPQHKLSLTASHOERLPST-----QELYAHG-----K 522
 Db 435 RAFFLQSGWKATDD--LTLASGVRQSQSMTDVSDFVAANOQLLIANGLGKTADAVPGSGK 492
 Qy 523 HVAITNTFEVGNKH-LMKERSNNIELALGYE--GDRWQYNLALRYRNFNGYIY-----AOTL 575
 Db 493 DIDVNLVNVGAIYKLNLOQOVWANYSEGPELDPKAYYGFGRYGAADGNHYPPLQGVSV 552
 Qy 576 NDGR-----GPKSIEDDSEMKLYRYNGSGADFYGA 605
 Db 553 NDSPLDGIKTKQVELGWRHTDQALDTQVAAFYSWSDKSIKYDSKTLAVLQOQTKRNYGL 612
 Qy 606 EGEIYEKPTPRYRIGVSGDYVRGLKMLPSLGR---EDAYGNRP-----FIA-QDDQNA 656
 Db 613 EGQATYWLDDHWQVGVNGLAIRSQEK---VDGRWLKQDVTASPSKAGAFVGMKDDQRS 668
 Qy 657 PRVPAARLGFHLKASLTDRIDANLDYRVF 686
 Db 669 LRQGVGR-----TFNLNDFPGNKIDGYALF 693

RESULT 27
 US-09-328-352-5942
 ; Sequence 5942, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 5942
 ; LENGTH: 761
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-5942

Query Match 4.3%; Score 174; DB 4; Length 761;
 Best Local Similarity 19.9%; Pred. No. 1.4e-07;
 Matches 180; Conservative 119; Mismatches 299; Indels 308; Gaps 46;
 Qy 7 KPIVLSIL---LINTPLLAQAHTESQ-----VGLTETVVVGKSRPRATSGL 50
 Db 10 KRIIQSVLSVSLASMMMAFAAQNEQDAEKTLEKPAEPVKLETIFVTAEQVKOSLGV 69
 Qy 51 LHTSTASDKIISGDTLRQKAV--NLGDALDGVPGIHASQVGGASAPVIRGQ----- 101
 Db 70 -----SVITQEDLEKLPVRNDISDYVRMPGVNLT---GNSATGQGNRQIDIRG 117
 Qy 102 -----GRIKVLN--HGETGDM-----ADSPDHAIWDTALSOQVEILRGV 143
 Db 118 MGPENTLILVDGKPIINSRVSRYGKGERDTRGDSNWPABAI-----ESIEVLRGPA 170
 Qy 144 TLLYSNGVAGLVADVADGKIPEK-----MPEN---GVSGELGLRLSS----- 182
 Db 171 AARYGSAAGGVNIIITKVTNETHSGVEFYTSQPEDSKESGNSRVGNFVSGPLIKDVL 230
 Qy 183 -----GNLEKLTSGGINI--GLGNFVLHTEGLYRK--SG-----DYAVPRY 220
 Db 231 YRLYNTNKTAEADDVDINKSIGST-AAGREGVKNKDISGLRAWQATDQQTLLADISSSQ 289
 Qy 221 RNLKRLPDSRRRANGQRAVLGWRKRFYRRTYDRDRDQYGLPAHSHEYDDCHADIWQK 280
 Db 290 GNI---YSGDSQLNANADAILS--QLIGKETNTIRDSYAL---THE-----GDSWKG 337
 Qy 281 SLINKRYLYQLPHLLTEEDVDYDNPGLSCGFHD-----DDAHAHNKPWIDLRKRY 335
 Db 338 SKLVAQYDKTHNKRLPE-----GLAGSVEKINNLLDDKATS-----RLETL 378


```
; SEQ ID NO 25815
; LENGTH: 595
; TYPE: prt
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25815

Query Match          4.2%; Score 169; DB 4; Length 595;
Best Local Similarity 20.8%; Pred. No. 2.8e-07;
Matches 159; Conservative 97; Mismatches 250; Indels 258; Gaps 42;

QY   78 DVPCHASQYGCGASAPVIRGQTGRRIKVLNHHGTGMADSPDHAIWDTALSQOQE 137
    |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   5 DGIP-----ASTPDGGQ-----|||||:|||||:|||||:|||||:
                                     |||:|||||:|||||:|||||:
QY   138 ILRGPTVTLLYSNGNVAGLVDP--ADGKIPEKMPENGVSVELGLRLSSGNLEKLTSGGINI 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   31 VLRGFPASTIYGS-NAGGVIQMFSDRGQGAPRVGAETVSGDG--LSRNHL--YTEGE--- 82
                                     |||:|||||:|||||:|||||:
QY   196 GLGKNFV----LHTEGLYKSGDYAVPR---YRNLKBLPOSPPRFANGQHRVILGWKR 246
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   83 GGVGVFLVDASRMOTDGYR--DHSAAARDQTFAKLNFRPADSRLA-----126
                                     |||:|||||:|||||:|||||:
QY   247 RFYRRYS-----DRRDQYGLPAHSHEYD-----DCHADIWQKSILIN-KRY-- 287
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   127 ----LIYSSLSEQNATEDPLGQTDWAYKHDPSPSVTANAELYDTRKSIDHQAGMANYERYFG 182
                                     |||:|||||:|||||:|||||:
QY   288 ---LOLYPHLLTBEDVDYNPGLSCGHDDDAHAHAGKPMIDLNRNKRYELRAEWKQP 344
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   183 EATLQWNAVYVKRSVVQYQAIPQIGC-----ESNPRCQRNGAVIDFDPDFHGTVRMQLP 238
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   345 F---PGFEALRVHLNRNDYHHDEKAGDAVENFNNTONARIELRHQPIGRUKGWSGVQY 401
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   239 VSAQAGEENLITVGLD-YDQSRDDRSG--YQN-FNGDLQGVKGELRDEVD-----284
                                     |||:|||||:|||||:|||||:
QY   402 LQOKSSALSATSBAVKOPMLLDNKKVQHYSFFGVEQANW--DNFTLBGGVRVKEQKASIRY 459
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   285 -----TATS---LDPYL-----QASWAIDAWTLQAGVRHSTMKEV---317
                                     |||:|||||:|||||:|||||:
QY   460 DKALIDRENYKYQLPDLGLAHROTARSFALSGNNYFTPQHKULSTASHOERLPSQOEL-Y 518
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   318 -----DDRYLSNGDSAGSRRYRKNTSPFSVMYAFTPDLLGHYISLAGKFETPTQAEAY 370
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   519 AHGKHVAINTFEVGNCHKLNKERSNNIELALCYEGCDRW---OYNLAYRNRFNGNIYAQTL 575
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   371 A---PVANAAPDVNFGLKPATSOYE--AGLKARLWGNTRVNAALFOJTEDEIVVASS 425
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   576 NDRGPKSIEDDSEMKLVRYNQSGADF-----YGAEGEI-----YFKTPTRYICV 621
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   426 LGCR-----TSYONAGKTLLRGFELGLSELSEHNWANLANLTRUSATYDSDF 472
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   622 -SGDYVRGLKNLRSLP-----GREDAYGNRPFIADQDONQAPRVPA 662
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   473 EAGGXTIKGKXHLPGVPRESSLFGBELVWKPAEGISMGWEQMYRSQVVV--EDSNSEKAAPS 530
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   663 RLGFHLKASLTDRIANDLDIRYRFAQNKLAARYETRPFCHMLNLGANYRRTYGENWYY 722
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   531 -----YAVF--NWRTRFE-----ORLGTMWFAP 550
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY   723 --VKADNLLINSVYAHSSFSDTPQM-----GRSFTGGVNVXF 758
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   551 OLVRIDNLNFDRO-YVGSVIVGDGNNRYVEEAAPGLSYVAGAGVY 593
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

RESULT 31
US-09-252-991A-30935
; Sequence 30935, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A

```



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,435
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-462 MIS.vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-478-435-5

```

```

Query Match 4.0%; Score 163; DB 2; Length 913;
Best Local Similarity 17.5%; Pred. No. 2.2e-06;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

Qy 5 TLKPI-----VLSILLINPLAQAH---ETEQSVG-----LETVTWVG-KSRP 44
Db 2 TKRPYFLRSLIISCLLISCVYKABTQSIKDTKEAISSEVDTQSTEDSELETISVTAEKVRD 61
Qy 45 RATSGLIHTSTASDKII--SGDTRQKAVNLGDALDGVFGIHASQYGGGASAPV-IRGQT 101
Db 62 RKDNEV---TGLGKIKTSESISRQVNLIRDLTRYDPGISVVEQGRGASSGYSIRGWD 117
Qy 102 GRIKVL-----NHCETGDMADFPDHALIMVDTALSQQVEILRGP 142
Db 118 RNRVALLVGLPQTQSVVQSVPLVARSYGSGTGAINIEYENV-----KAVSISKGG 169
Qy 143 VTLIYSSGNVAGLV-----DVADG-----KIPEKMPENGVSGBELGLRLSSGNLE 186
Db 170 SSSEYGNALGAGSVTFQSKSAADILEGDKSWGIGTQKNAYSKNKGFTHS LAVAGKGGFE 229
Qy 187 KLISGGINIGLGN-----FVLHTE--GLYRKSQDYAVPRNLKR 225
Db 230 -----GVAIYTHRNSTIQTQVHKDALKGVOSYDRFIATTDQSAYFVMQDECLDGYDKCKT 284
Qy 226 LPDSRFRFANGQRAVL-----GWRKRFRYRTYSRRDQY--GLPA 264
Db 285 SPKEPATLSQRETVSVSDYTGANRIKPNKPKYESQSW---FLRGYHSEHQHYIGGI--- 339
Qy 265 HSHYDDCHADIIWQKSLINKRVLQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHANG 324
Db 340 --FEFTQCKFDI-----RDMTPPAYLEPTEDKDLQ-----SRFPYKQDYGAVQHIG 384

```

```

RESULT 35
US-08-337-483-5
Sequence 5, Application US/08337483
Patent No. 5923562
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Hartness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-337-483-5

Query Match 4.0%; Score 163; DB 2; Length 913;
Best Local Similarity 17.5%; Pred. No. 2.2e-06;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

QY 5 TLKPI---VLSILLINTPLLAQAH---ETEQSVG-----LETVTIVVG-KSRP 44
Db 2 TKKPYFRLSIISCLLSYCYVKAETQSIKDKEALISEVDFTQSTEDSELETISVTAEKVRD 61
QY 45 RATSGLHTSTASDKII--SGDTLRQKAVNLGALDGVPGIHASQYGGGASAPV--IRGT 101
Db 62 RKDNEV---TGLGKIITKTSIESISREQVLNIRDLTRYDPGISVVEQGRGASSGYSIRGMD 117
QY 102 GRRIKVL-----NHHGETGDMADFSDPDAIMVDTALSQOQVEILRGP 142
Db 118 RNRVALLVDGLPQTQSYVQSPVARSYSYSGTGAINIEYENV-----KAVEISKGG 169
QY 143 VTLYSSGNVAGLV-----DVADG-----KIPEKMPENGVSGBELGLRLSGNLE 186
Db 170 SSSEYGNALAGSVTFQSKSAADILEGDKSWGIGTQKNAYSCKNGKFTHSLAVAGKQGGFE 229
QY 187 KLTSGGINIGLGNK-----FVLHTE--GLYKSGDYAVPYRNLKR 225
Db 230 -----GVAIYTHRSIETQVHKDALKGVQSVDFRIATTEQSAFYVQWDECLDGYDKCT 284
QY 226 LPDSRRFPANQRAVL-----GWRKRFYRTYSRRDQY--GLPA 264
Db 285 SPKRPATLSTQRETQSVSDYTGANKRIPKPMKYESQSW---FLRGGVHFSEQHYIGI-- 339
QY 265 HSHEVDDCHADIIWOKSLINKRYLQLYPLHLLTEDVDYDNPGLSCGFPHDDDDAHAAHNG 324
Db 340 --PFTQKQFDI-----RDMTFPAYLRPTEDKDLQ-----SRFPYKQDYGVQCHIG 384
QY 325 KPWIDLNRKRYELRAEWKQPPGFPEALRVHLNRNDYHDEKAGDAVENFNNQTNQARIE 384
Db 385 ----DGRGVKY-----ASGLYFDEHHRKORVG--IYIYENKNAKAGIID 422
QY 385 LRHQPIGRKLSGWSGVQYLGQSSALSATSEAVKQPMILLDNKVQH----- 428
Db 423 -----KAVLSANQQNIILDSYMRHTHCSLYPNFSKNCRPT 457
QY 429 ---YSFPGVE-----QANW-----DNFT----- 443
Db 458 LDKPYSYTHSRNRYVYKEKHNMLQNLLEKTIQQNWLTHQIAFNIGFDFTSALQHDYLTR 517
QY 444 --LEGGVVEKOKASIRYDKALIDRENYVKOPLP-----DLCAHROTARSA----- 488
Db 518 RVIATASSISEKRGAEARNGL---QSSPYLYPTPKAELVGGDLQNYQKSSNYSDCKVRL 574
QY 489 LSG-NWYTPQHKULSLTASHQERLPSTQELYAHQKHVATNTPFVGNKH-LNKRSNNIEL 546
Db 575 INKKNYFAARNMAL-----GKYV---DLGLGMYDVSRTKANESTI 614
QY 547 ALG-YEGRWQVNLALYRNRFNGNYIAQTLDNG-RGPKSIEDDSEMKLVYNQSGADFYG 604
Db 615 SVGKEKFNFSWNTGVIKTEWMLDSY--RLSTGRNPSF-----AEMYGWYGGKDTVIYI 668
QY 605 ABGEIYFKP-----TPRYIGVSGDYVRGRNLKNLSLPGREDAYGNRPFIAQDQONAPRVP 660
Db 669 GK-----FKPETSRNQEFGLALKGDFGNTIEISHF-----SNAYRNLIATFAEELSCKNGTTG 718

```

```

QY 661 AARLGFH-----LKASLTDRIDAN-----LDYRVVFAQNKLA----- 692
Db 719 KENYGVHNAQNAKLVGVNITAOQLDFRGLMKRIPYGYWYATFAYNRVVKYDKQKINAGLASVS 778
QY 693 --RYETRTPGHMLNLGANYRRNTRYGEMWVYVKADNLLNQSVYAHSSFLSDTPQMGSRSF 750
Db 779 SYLFDATQPSRYIIGLYDHPSTN-----WGIK-----TWFTQSKAKSQNELLGKRA 825
QY 751 TG--GVNVK 757
Db 826 LGNNSRNVK 834

```

```

RESULT 36
US-08-478-373-5
; Sequence 5, Application US/08478373
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-373-5

```

```

Query Match 4.0%; Score 163; DB 2; Length 913;
Best Local Similarity 17.5%; Pred. No. 2.2e-06;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

```

```

QY 5 TLKPI-----VLSILLINTPLLAQAH---ETQSVG-----LETVTVG-KSRP 44
Db 2 TKKPYRLSIISCLLSICVKAETQSIKDTKEAISSEVDTSQSTSELETISVTAERV 61
QY 45 RATSGLHTSTASDKII--SGDTLRQKAVNLGDALDGVPGIHASQYGGASAPV-IRGOT 101
Db 62 RKDNEV---TGLGKIITSEISREQVNLNRLTRYDPGISVVEQGGASSGYSIRGMD 117
QY 102 GRIKVL-----NHHGETGDMADFPDHAIMVDTALSOQVEILRGP 142
Db 118 RNRVALLVDGLPOTQSVVQSPLVARSYSGTGAINIEYENV-----KAVEISKGG 169
QY 143 VLLYSSGNVAGLV-----DVADG-----KIPEKPENGVSGELGRLSSGNLE 186
Db 170 SSSEYNGNLAGSVTFQSKSAADILEGDKSWGIQTKNAYSSKNKGFTHSLAVAGKQGF 229
QY 187 KLTSGGINIGLGN-----FVLHTE--GLYKSGDYAVPYRNLKR 225
Db 230 -----GVAIYTHNSIETQVHKDALKGVQSYDRFIATTTEDQSAYFVMQDECLDYDKCT 284
QY 226 LPDSPRRFANGQRAVL-----GWRKRYRRTYSDDRDQY--GLPA 264
Db 285 SPRKPTLSQRTSVSVSYDTGANRIKPNPMKYESQSW---FLRGYHFSQHYIGI-- 339
QY 265 HSHEYDDCHADIIWQKSLINKRYQLYPHLLTEEDVDNPNGLSCGPHDDDAHAHANG 324
Db 340 --FEFTQOKFDI-----RDMTPPAYLRPTEDKDLQ-----SRFPYKQDYGAYQHIG 384
QY 325 KPWIDLRNKEYELRAEWKQFPFGFEALRVHLNRNDVHDEKAGDAVENFFNNOCTNARIE 384
Db 385 ---DGRGVY-----ASGLFDEHRRKQVVG--JEIYENKVKAGILD 422
QY 385 LRHPQIGRLKXGNGVQYLOKSSALSATSEAVKQPMLLDNKVOH----- 428
Db 423 -----KAVLSAQONILDSYMRHTHCSLYPNPSKNCRPT 457
QY 429 ---YSFEGVE-----QANW-----DNFT----- 443
Db 458 LDKPYSTHSRNVYKKNMLQNLKLEKIQQNLWTHQAFNLGDFDFTSALQHKDYLTR 517
QY 444 --LEGGVRVBEQKASIRYKALIDRENYKQPLP-----DLGAHROTARSAFA----- 488
Db 518 RVATASSISSEKGEARRNGL---QSSPYLPTPKAELVGGDLNCGYQKSSNYSCKVRL 574
QY 489 LSG-NWYFTTQHKLSLTASHOERLPSTQELYAHKHAATNTEVGNKH-LNKERNNIEL 546
Db 575 IKGNVYFAARNWAL-----GKIV---DLGLGNRYDVSKYANESTI 614
QY 547 ALG-YEGDRWQYNLALYRNFQNYVIAQTLDNG-RGPKSIEDDSMKLVRYNQSGADPYG 604
Db 615 SVGKFKNFSWNTGIVIKPTEWLDLSY--RLSTGFRNPSE---AEMYGWRYGGKDTDVI 668
QY 605 AEGBIYFKP---TPRYRIGVSGDYVRGLNKLPSLPGRDQVGNRPFFTAQDDQNAVRP 660
Db 669 GK---PKPETSQNEFGLAKGDFGNIEISHE-----SNAYNLLIAPABELSKNGITG 718
QY 661 AARLGPH-----LKASLTDTRIDAN-----LDYRVFAQNKLA----- 692
Db 719 KGNYGYNHQAQNAKLGVNITAQDLDFNLGKRIPIYGYATFAYNRVVKDQKINAGLASVS 778
QY 693 --RYETRTPGHMLNLGNAVRNRYGEMWNVYKADNLLNQSYYAHSSFLSDTPQMGRSF 750
Db 779 SYLFDATQPSRYIIGLYDHPSTI-----WGIK-----TWFTQSKAKSQNELLGKEA 825
QY 751 TG--GVNVK 757
Db 826 LGNNSRVK 834

```

RESULT 37
US-08-474-671-5
; Sequence 5, Application US/08474671

```

; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mardin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESS: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MIS.vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-474-671-5

```

```

Query Match 4.0%; Score 163; DB 3; Length 913;
Best Local Similarity 17.5%; Pred. No. 2.2e-06;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

Qy 5 TLKPI-----VLSILLINTPLLAQAH---ETQSVG-----LETVTVG-KSRP 44
Db 2 TKKPYRLSIISCLLSICVKAETQSIKDTKEAISSEVDTSQSTSELETISVTAERV 61
Qy 45 RATSGLHTSTASDKII--SGDTLRQKAVNLGDALDGVPGIHASQYGGASAPV-IRGOT 101
Db 62 RKDNEV---TGLGKIITSEISREQVNLNRLTRYDPGISVVEQGGASSGYSIRGMD 117
Qy 102 GRIKVL-----NHHGETGDMADFPDHAIMVDTALSOQVEILRGP 142
Db 118 RNRVALLVDGLPOTQSVVQSPLVARSYSGTGAINIEYENV-----KAVEISKGG 169
Qy 143 VLLYSSGNVAGLV-----DVADG-----KIPEKPENGVSGELGRLSSGNLE 186
Db 170 SSSEYNGNLAGSVTFQSKSAADILEGDKSWGIQTKNAYSSKNKGFTHSLAVAGKQGF 229

```



```

Db 423 -----KAVLSANQONILDSYMRHCHSLYPNPSKNCRPT 457
Qy 429 ----YSFGVE-----QANW-----DNFT-----443
Db 458 LDKPSYSHSDRNVYKEKHNMQLNLEKKIQONWLTHQIAFNGLGDFDFTSALQHKDYLTR 517
Qy 444 --LEGVRVEKOKASIRYDKALIDRENYKQPLP-----DLGAHQRTARSA-----488
Db 518 RVATASISEKRGARNGL---OSSPYLYPTKAEVLGGLDLCNYGKSSNYSQCKVRL 574
Qy 489 LSG-NWYFTPHQKLSLTASHOERLPSTQELVYAHGKHVATNTFEVGNKH-LNKERSNNIEL 546
Db 575 IKGNYYFAARNMAL-----GKYV---DLGLGMYDVSRTKANESTI 614
Qy 547 ALG-YEGDRWQYNLALYRFRGNYIYAQTLDNG-RGPKSIEDDSEMKLVYNSQGFY 604
Db 615 SVGKFKNSWNTGVIKTEWLDLSY--RLSTGFRNFSF---AEMYGWYRGKDTDYI 668
Qy 605 ABEETVFKP---TPRYRIGVSGDYVRGRKLNPLSLFGREDAYGNRRFFIAQDDQNA PRVP 660
Db 669 GK---FKPETSROEFGALKGDFGNIESHF-----SNAYRNLIATAFAEELSKNCTTG 718
Qy 661 AARLGFH-----LKASLTBRIDAN-----LDYRVFPAOKLA-----692
Db 719 KGNYYHNAQNAKLVGVNITAQDLFNGLWKRIPIYGVWATFAYNRVYKVDQKINAGLASVS 778
Qy 693 --RYETRTPGHMLMGANRYRNTRYGEWNYVAKADNLLNQSVYAHSSFLSDTPQMGRSF 750
Db 779 SYLFDALQPSRYIIGLYDHPST-----NGIK-----TMTQSKAKSQNELLGKRA 825
Qy 751 TG-GVNUK 757
Db 826 LGNRSNRX 834

```

RESULT 39
 US-08-897-438-5
 ; Sequence 5, Application US/08897438
 ; Patent No. 6282016
 ; GENERAL INFORMATION:
 ; APPLICANT: Loomis, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 160
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/897,438
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/483,577
 ; FILING DATE: 07-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/337,483
 ; FILING DATE: 08-NOV-1994
 ; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-897-438-5

```

Query Match 4.0%; Score 163; DB 3; Length 913;
 Best Local Similarity 17.5%; Pred No 2.2e-06;
 Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

```

Qy 5 TLKPI---VLSILLINTPLLAQH---ETQSVG-----LETVVVG-KSRP 44
Db 2 TKKPYFRLSIISCLISVCVKAETQSIKDKAEISSEVDQTQSTEDSELETISVTAERVD 61
Qy 45 RATSGLHTSASDKII--SGDTLRKAVNLGDALDVGPIHSAQYGGASAPV-IRGQT 101
Db 62 RKNVE---TGLGKIYTESISREQLNTRDTRDPLGISVVEQGRGASSGYSIRGMD 117
Qy 102 GERIKV-----NHHGTGMDADPSPDPAHIMVPTALSSQQVEILRGP 142
Db 118 RNRVALLVDGLPQTQSVYVQSPLVARSYSGTGAINEIYENV-----KAVEISKGG 169
Qy 143 VTLLYSSGNVAGLV-----DVADG-----KIPEKMPENGVSGLRLSSGNLE 186
Db 170 SSSEYNGALAGSVTFQSKSAADILEGDKSWGIOQTKNAYSSKNKGFTSHSLAVAKQGGFE 229
Qy 187 KLTSGGINIGLKN-----FVLHTE--GLYKRGSDYAVPRYRLMR 225
Db 230 ----GVALYTHRSIETQVHKDALKGVSYDRFTATTEDQSAFVVMQDECLDGVKCKT 284
Qy 226 LPDSRFRFANGQRAVL-----GWRKBFYRTYSDRDOY--GLPA 264
Db 285 SPKRPTLSTQRETQSVSDYTGANRIKPNMKYESQSW---FLRGYHFSHQYIGGI-- 339
Qy 265 HSHEDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNFGLSCGPHDDDDAHAAHNG 324
Db 340 --FEFTQCKFDI-----RDMTFPAYLRPTEDKDLQ---SRPFYPKQDYGAYQHIG 384
Qy 325 KPWIDLNRKRYELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFNNQTONARIE 384
Db 385 ----DGRGVKY-----ASGLYFDHHRKQRVG--IEIYIENKKNAGIID 422
Qy 385 LRHQPIGRKLSGWQVYLGQSKSALSATSEAVKQPMILDKNVQH-----428
Db 423 -----KAVLSANQONILDSYMRHCHSLYPNPSKNCRPT 457
Qy 429 ----YSFGVE-----QANW-----DNFT-----443
Db 458 LDKPSYSHSDRNVYKEKHNMQLNLEKKIQONWLTHQIAFNGLGDFDFTSALQHKDYLTR 517
Qy 444 --LEGVRVEKOKASIRYDKALIDRENYKQPLP-----DLGAHQRTARSA-----488
Db 518 RVATASISEKRGARNGL---OSSPYLYPTKAEVLGGLDLCNYGKSSNYSQCKVRL 574
Qy 489 LSG-NWYFTPHQKLSLTASHOERLPSTQELVYAHGKHVATNTFEVGNKH-LNKERSNNIEL 546
Db 575 IKGNYYFAARNMAL-----GKYV---DLGLGMYDVSRTKANESTI 614
Qy 547 ALG-YEGDRWQYNLALYRFRGNYIYAQTLDNG-RGPKSIEDDSEMKLVYNSQGFY 604

```

```

Db 615 SVGKFNFSWNTGIVIKPTWLDLSY--RLSTGFNPSF---ARMYGRYCGKDTDVI 668
Qy 605 ABGEIYFKP---TPRYRIGVSDYVRGLKNLPSLPGREDAYGNRPFIADDDQNAVRP 660
Db 669 GK-----FKPETSQNFGLKGDGFGNIEISHP-----SNAYRNLIAPABELSKNGTGTG 718
Qy 661 AARLGFH-----LKASLTDRIDAN-----LDYRVFAQNKLA----- 692
Db 719 KGNVGHNAQNAKLVGNVITAQDLDFNGLMKRIPYGYATFAYNRVKVDQKINAGLASVS 778
Qy 693 --RYETTPCHHMLNIGANVRNTRYGEWNWYKADNLLNQSVYAHSSFLSDTPQMGRSF 750
Db 779 SYLFDALQPSRYIIGLYDHPSTNT-----WGIK-----TMTQSKAKSONELLGKRA 825
Qy 751 TG--GVNVK 757
Db 826 LGNSNRNVK 834

```

RESULT 40

```

US-08-637-654-5
; Sequence 5, Application US/08637654
; Patent No. 6358727
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Harkness, Robin E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew D
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA94/00616
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-595
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-637-654-5

```

Query Match 4.0%; Score 163; DB 4; Length 913;
 Best Local Similarity 17.5%; Pred. No. 2.2e-06;
 Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

```

Qy 5 TLKPI-----VLSTLLINTPLLAQAH-----ETESVSG-----LETVTWVG-KSRP 44
Db 2 TKKPYFRLSTIISCLISCYVKAETQSIKDKAEISSEVDSTQSTEDSELETISVTAEKVRD 61
Qy 45 RATSGLLHSTASDKII--SGDTLRQKAVNLGDALDVGPGIHASOVGGGASAPV-IRGT 101
Db 62 RKNFV-----TGLGKIITKTSISISREQVNIIRDLTRYDPGISVVEQGRGASSYIRGMD 117
Qy 102 GRRIKVL-----NHHGETGDMADFPDPHAIMVDTALSOQVILRGP 142
Db 118 RNRVALLDCLPQTQSVYVQSPLVARSYSGTGCAINEIYENV-----KAVEISKGG 169
Qy 143 VTLLYSSGNVAGLV-----DVADG-----KIPEKPENGVSGELGLRSSGNLE 186
Db 170 SSSEYNGGALAGSVTPQSKSAADILEGDKSWGIQTNKAYSSKNKGFTTHSLVAGKQGF 229
Qy 187 KLTSGGINIGLGNK-----FVLHTE--GLYRKSGDYAVPRYRNLR 225
Db 230 -----GVAIVTHRNSIETQVHKDALKGVQSYDRFTATTEDQSAFYVMDCLDGYDKCKT 284
Qy 226 LPDSPPRFANGQHRVL-----GWRKRFYRYRTYSRRDOY--GLPA 264
Db 285 SPKRPATLSTQRETVSVDYTGANRIKPNMKYESQSM--FLRGYHFSEQHYLGGI-- 339
Qy 265 HSHEYDDCHADIIWQKSLINKRYLQLYPHLLTTEEDVDYDNPLGSCOFFHDDDDAHAHNG 324
Db 340 --FEFTQOKFDI-----RDMTPPAYLRPTDKDLQ-----SRPFYKQDYGAYQHIG 384
Qy 325 KPWIDLNKYEYELRAEWKQFPFGFALRVHLNRNDYHDEKAGDAVENFNNTQONARIE 384
Db 385 ---DGRGVKY-----ASGLVFDEHHRKQRVG--TEYIYENKNKAGI 422
Qy 385 LRHQPIGRLKSGWGVQYLGQKGSALSATSEAVKQPMLLDNKVQH----- 428
Db 423 -----KAVLSANQONILDSYMRHTHCSLYPNPKNCRPT 457
Qy 429 ---YSPFGVE-----QANW-----DNFT----- 443
Db 458 LDKFYSYHSDRNVYKEKHNMLQLNLEKKIQONWLTHQIAFNLPDFDFTSALQHKDYLTR 517
Qy 444 --LEGGVREVKQKASIRYDKALIDRENYKQPLP-----DLGAHQRTARSAF----- 488
Db 518 RVITATASSISEKGEARNGL---QSPVLYPTPKAELVGGDLQVYQKSSNVSDCKVRL 574
Qy 489 LSG-NWYFTPOHKLSLTASHQERLPSTQELYAHGKHVATNTEVGNKH-LNKERSNNIEL 546
Db 575 IKGNYYFAARNMAL-----GKYV---DLGLGMRDYVSRTKANESTI 614
Qy 547 ALG-YEGDRWQYNLALYRNRFNGYVYAQTLDNG-RGPKSIEDDSEMKLVRYNQSGADFYG 604
Db 615 SVGKFNFSWNTGIVIKPTWLDLSY--RLSTGFNPSF---ARMYGRYCGKDTDVI 668
Qy 605 ABGEIYFKP---TPRYRIGVSDYVRGLKNLPSLPGREDAYGNRPFIADDDQNAVRP 660
Db 669 GK-----FKPETSQNFGLKGDGFGNIEISHP-----SNAYRNLIAPABELSKNGTGTG 718
Qy 661 AARLGFH-----LKASLTDRIDAN-----LDYRVFAQNKLA----- 692
Db 719 KGNVGHNAQNAKLVGNVITAQDLDFNGLMKRIPYGYATFAYNRVKVDQKINAGLASVS 778
Qy 693 --RYETTPCHHMLNIGANVRNTRYGEWNWYKADNLLNQSVYAHSSFLSDTPQMGRSF 750
Db 779 SYLFDALQPSRYIIGLYDHPSTNT-----WGIK-----TMTQSKAKSONELLGKRA 825
Qy 751 TG--GVNVK 757
Db 826 LGNSNRNVK 834

```

Search completed: November 14, 2003, 11:01:15
 Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:53:22 ; Search time 17 Seconds

(without alignments)

2096.838 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLKPIVLISILLINTPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	231	5.7	652	1	IRGA_VIBCH
2	228	5.6	744	1	HXC1_HAEIN
3	221	5.5	723	1	X262_HAEIN
4	211	5.2	673	1	FYUA_YERPE
5	210	5.2	673	1	FYUA_YERPE
6	205.5	5.1	676	1	HMUR_YERPE
7	203	5.0	725	1	HXC2_HAEIN
8	187.5	4.6	663	1	CIRA_ECOLI
9	186.5	4.6	687	1	HEWR_YEREN
10	184.5	4.6	614	1	BTUB_SALTY
11	184.5	4.6	810	1	HPUB_NEIMC
12	176	4.4	614	1	YNGD_ECOLI
13	174.5	4.3	700	1	HPUB_NEIMA
14	169.5	4.2	810	1	RHTA_RHIME
15	161	4.0	746	1	FPEA_ECOLI
16	160.5	4.0	746	1	FBEL_ECOLI
17	158.5	3.9	912	1	TFB1_NEIMB
18	155.5	3.9	746	1	PFBA_PSEAE
19	155	3.8	908	1	IBT2_NEIMB
20	154	3.8	732	1	IUTA_ECOLI
21	151.5	3.8	710	1	FOXA_YEREN
22	150.5	3.7	915	1	TBPI_NEIGO
23	144.5	3.6	911	1	TBPI_NEIME
24	141.5	3.5	729	1	FHUE_ECOLI
25	139	3.4	702	1	FOXA_SALTY
26	138.5	3.4	943	1	LBPA_NEIMB
27	134.5	3.3	720	1	FPTA_PSEAE
28	133	3.3	809	1	PUPB_PSEPU
29	131	3.2	635	1	HPUC_VIBCH
30	130	3.2	819	1	PUPA_PSEPU
31	129	3.2	1476	1	GTFE_STRMU
32	127.5	3.1	944	1	LBPA_NEIMA
33	127	3.1	815	1	FPVA_PSEAE

RESULT 1

ID	IRGA_VIBCH	STANDARD;	PRT;	652 AA.
AC	P27772; Q9KUP0;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Iron-regulated outer membrane virulence protein precursor.			
GN	IRGA OR VC0475.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;			
OC	Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=666;			
RN	[1] _			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;			
RX	MEDLINE=93023868; PubMed=1406279;			
RA	Goldberg M.B., Boyko S.A., Butterton J.R., Stoeber J.A.,			
RA	Payne S.M., Calderwood S.B.;			
RT	"Characterization of a Vibrio cholerae virulence factor homologous to			
RL	the family of TonB-dependent proteins.";			
RL	Mol. Microbiol. 6:2407-2418(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=E1 Tor N16961 / Serotype O1;			
RX	MEDLINE=20406833; PubMed=10952301;			
RA	Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,			
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,			
RA	Gill S.R., Nelson K.R., Read T.D., Tettelin H., Richardson D.,			
RA	Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,			
RA	McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,			
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,			
RA	Fraser C.M.;			
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio			
RL	cholerae.";			
RL	Nature 406:477-483(2000).			
RN	[3]			
RP	SEQUENCE OF 1-151 FROM N.A.			
RC	STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;			
RX	MEDLINE=91072235; PubMed=2174861;			
RA	Goldberg M.B., Boyko S.A., Calderwood S.B.;			
RT	"Transcriptional regulation by iron of a Vibrio cholerae virulence			
RL	gene and homology of the gene to the Escherichia coli fur system.";			
RL	J. Bacteriol. 172:6863-6870(1990)			
CC	-1- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING			
CC	FERRIC VIBRIOBRADIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS			
CC	V. CHOLERAEE TO EXTRACT IRON FROM THE ENVIRONMENT.			
CC	-1- SUBCELLULAR LOCATION: Outer membrane.			
CC	-1- MISCELLANEOUS: REGULATION OF THE IRGA EXPRESSION IS NEGATIVELY			
CC	REGULATED AT THE TRANSCRIPTIONAL LEVEL BY IRON			
CC	-1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

34	123.5	3.1	624	1	HTPG_SALTY	P58480 salmonella
35	122.5	3.0	624	1	HTPG_SALTY	P58479 salmonella
36	122	3.0	634	1	HTPG_VIBPA	Q87rh5 vibrio para
37	120.5	3.0	713	1	AMVR_BACS8	P1692 bacillus sp
38	120.5	3.0	790	1	YDDB_ECOLI	P1827 escherichia
39	119.5	3.0	730	1	GELS_HORSE	Q28372 equus cabal
40	119.5	3.0	913	1	YC17_HAEIN	P45114 haemophilus
41	119.5	3.0	1067	1	HGBB_HAEIN	Q9kiv1 haemophilus
42	119	2.9	622	1	HTPG_YERPE	P58482 versinia pe
43	118.5	2.9	712	1	CDGT_BACS3	P09121 bacillus sp
44	118.5	2.9	713	1	CDGT_BACS0	P05618 bacillus sp
45	118.5	2.9	718	1	CDGT_BACCI	P30920 bacillus ci

ALIGNMENTS

Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 NCBI_TaxID=727;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RC MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kierulff A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
 RA McKenny K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RT Science 269:496-512 (1995).
 RN [2]
 RN IDENTIFICATION BY MASS SPECTROMETRY.
 RC MEDLINE=20137488; PubMed=10675023;
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wigf B.,
 RA Gray C., Fountoulakis M.;
 RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
 RT Electrophoresis 21:411-423 (2000).
 CC -|- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
 CC -|- SUBCELLULAR LOCATION: Outer membrane (potential).
 CC -|- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).
 CC -----
 CC EMBL; U32712; AAC21927.1; --
 DR PIR; C64058; C64058.
 DR TIGR; H10262;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 DR Outer membrane; Receptor; Signal; TonB box; Complete proteome.
 KW SIGNAL 1 21
 FT CHAIN 22 723 PROBABLE TONB-DEPENDENT RECEPTOR HI0262.
 FT SITE 706 723 TONB C-TERMINAL BOX.
 FT SITE 723 AA; 80775 MW; 91EB3AB0FFEA2984 CRC64;
 SQ SEQUENCE 723 AA; 80775 MW; 91EB3AB0FFEA2984 CRC64;
 Query Match 5.5%; Score 221; DB 1; Length 723;
 Best Local Similarity 19.6%; Pred. No. 1.2e-07;
 Matches 157; Conservative 130; Mismatches 301; Indels 212; Gaps 41;
 Qy 11 LSLAILTPLLAQAHETEQVGLTEVTVVGKSRPRATSGLLTSTASDKIISGDTLRKA 70
 Db 6 LSLAIITTT-LVTANALAQSVELDSINVIATRP---SRFAYTPEQSK---DSLSSKA 57
 Qy 71 VNLGALGVGFIHASQVGG---ASAPVIRGQTGR-IKVLNHGEGTDMADPSDH--- 124
 Db 58 TSVAADLEIDIPNDVR--GGSRSTAQKPNIRIGSDNRVVDIGVRQNFDA----HRG 110
 Qy 125 ALMVDTALSQQVEILRGPTVTLIYSSGNVAGLVDAVDCKIPKMPNGVSGELGLRLSSG- 183
 Db 111 SYFLPMSLIQEIVIKGPFSSSLMGSGALGVVAVRTPNALDLLKNN---DKFGVKIRGY 167
 Qy 184 -NLEKLTSGGGINI-GLGNKVFVLTGELYKRSQDYAVPRVNLKRLPDSPPRFANGQHRV 241
 Db 168 QTANLSEKDVSVFAANDKFDVLISGPNYADNLRTGKGNKLN--TAYKQFGG---LAK 222
 Qy 242 LGWRKRFVRRY--YSDRDQYGLPAHSHEYDDCHADIIWOKSLINKRYLLYPLHLESDV 300

Db 223 FGQIINDANRVELSHRETRFKQTAPSN--EVENELTNEQIT 262
 Qy 301 DYDNPGLSCGFHDDDAHAHNGKWPIDLRNKRVELRAEMKQPPG-----FEALRVHL 355
 Db 263 DQTKK-----FHQCKD-----DLLPPTQPSERSSEFYKVKTRL 298
 Qy 356 NRNDYHHDEKAGD--AVENFF---NNQTONARIELRHQFGLKSGWGVYLGQKSSALS 410
 Db 299 GSVSYLTDOQIPQOSTVFNFYLPDPYLNTHIALYNNKT-----IERQKRVGVK 350
 Qy 411 ATSEAVKQPMLLDN--KVQHYSP-FGVEQANNDNFTLEGVVRVEKQ-----454
 Db 351 DQTKLTRGINLNSSELSHISFVYGVDMR-DKTRTERGTGSDAKFRADPYANSNWT 409
 Qy 455 -----ASIRYD-----KALIDRENYKOPLPDLGAHQRTARSFALS 490
 Db 410 GUYLIAHIFLFGKELVSPSVRYDHYDTGSKTVKYKDNHL-----SPATK 454
 Qy 491 GNVYFTPOHKLSTASHQE--RLPSTOELVYAGKHVATWTFE-----VGNKHLNERS 541
 Db 455 LTWIVT--NWLDFTAKYNEAFRAPSQERFVSGAHFGANTLGLDHNRFVANPNLRPETA 512
 Qy 542 NNIEL-----ALGVEGRWQYNLALYRNRFNYIYAQTLND-----GRGPKS-----583
 Db 513 KKEITANLHFDLSLFQGGKFKIEATYFRNDVKDFINLKIFNDAKTSASAGANPNTGAL 572
 Qy 584 LEDDSEMKLVRYNSGADFVGAEGEYVEKPTPRYRIGVSGDYVRGLKNLPSLPGREDAY 643
 Db 573 LPKNSQVQ-----NITWALSGIELQAOYQ--TEELTL-----FTNYGSKDKDQS 617
 Qy 644 GNRFFIAQDDQONAPRVPVPAARLGHKLKASLTD---RIDANLDYRVFAQNKLYETRT-P 699
 Db 618 G-----EALSNIASAASIGVGVNVALYKOKFTVGATWTHYA--AQRVVPKDHSTVTP 666
 Qy 700 GHMLNLGANYRNTRYGEW 719
 Db 667 SYILTDURAYYA--PLKGEW 684

RESULT 4
 FYUA_YERPE STANDARD; PRT; 673 AA.
 AC P46359;
 DT 01-NOV-1995 (Rel. 22, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Pesticin receptor precursor (IRPC).
 GN FYUA OR PSN OR YPO1906 OR Y2404.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=67/69;
 RA Rakin A., Heesemann J.;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM6;
 RA Fetherston J.D., Lillard J.W., Perry R.D.;
 RT "Analysis of the pesticin receptor from Yersinia pestis: role in
 iron-deficient growth and possible regulation by its siderophore.";
 RL J. Bacteriol. 177:1824-1833(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=67/69;
 RA Buchrieser C., Rusniok C., Couve E., Frangeul L., Billault A.,
 RA Kunst F., Carniel E., Glaser P.;
 RT "DNA sequence of the 102 kbases unstable region of Yersinia pestis.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]

RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RA MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titchall R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Irarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 RA MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Petherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
 CC SIDEROPHORE YERSINIABACTIN.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- INDUCTION: BY IRON STARVATION. IRON REGULATION MEDIATED THROUGH
 CC THE FUR PROTEIN.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z35104; CAA84487.1; -;
 CC EMBL; U09530; AAA69906.1; -;
 CC EMBL; AL031866; CAA21395.1; -;
 CC EMBL; AJ414150; CAC90722.1; -;
 CC EMBL; AE013845; AAM85962.1; -;
 CC PIR; A56148; A56148.
 CC PIR; AF0232; AF0232.
 CC HSSP; P05825; IREP.
 CC InterPro; IPR000531; TonB_boxC.
 CC Pfam; PF00593; TonB_dep_Rec; 1.
 CC PROSITE; PS00430; TONB DEPENDENT REC_1; FALSE NEG.
 CC PROSITE; PS01156; TONB DEPENDENT REC_2; FALSE NEG.
 CC Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
 CC Transp.; TonB box; Complete proteome.
 CC SIGNAL 1 22 POTENTIAL.
 CC FT CHAIN 23 673 PESTICIN RECEPTOR.
 CC FT SITE 30 37 TONB BOX.
 CC FT SITE 657 673 TONB C-TERMINAL BOX.
 CC FT CONFLICT 484 484 G -> D (IN REF. 1).
 CC FT CONFLICT 514 514 R -> G (IN REF. 1).
 CC SEQUENCE 673 AA; 73782 MW; 9C39E6010EBCE2C CRC64;
 Qy Query Match 5.2%; Score 211; DB 1; Length 673;
 Db Best Local Similarity 21.8%; Pred. No. 5.3e-07;
 Qy Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;
 Db 1 MAOTLKPIVLSILLINTPLAQAHTEQSVGLETVVVKSRPRATSGLLHTSTASDKI 60
 Qy 1 MKMTRLYPLALGGLL--FAIANAQTSQDD---ESTLVVTASKQSSRSA--SANNVSTV 53
 Qy 61 ISGDTLQKAVNLGDAIDGV-PGIHASQYGGGASAPV-IRGQTGRRIKVLNHHGETGMA 118
 Db 54 VSAPELSDAGVTASDKLPVLPGLNENSGNMLFSTISLRGVSSAQ-----99
 Qy 119 DF-SPDHAINVD-----TALS--QQVEILRGPVTLVSSGNVAGLVADVADGKIPE 165

```

Db 100 DFNPAVTLVYDGVGPOLSTNTIQAULTDQSVELLRGPGQTLGKSAQGGIINIVTQO-PD 158
QY 166 KMP-----ENGVSCELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRK--SGDYAVPR 219
Db 159 STPRGVIEGGVSSRDSYR-SKFNL-----SGFIQDGL-----LYGSVTLRLQVDDGDMINPA 209
QY 220 YRN-----LKRLPD-----SPRRFANGQHRAVLGWRKRFYRTYSDR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPMWGMFAASRECTRATQDAYVGNW-----258
QY 257 RDQYGLPAHSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDD 316
Db 259 -----DIKGRKLSISDGSPPDPMRRCT-----DSQTLSGKYTTDD- 293
QY 317 AHAHAHNGKFWIDLRNKRKYELAEWKQ-----PPGFPAALRVHLNRNDYHHDEK----- 365
Db 294 -----WV-----FNLSAWQOQHYSTFPF9--GSLIVNMPQR--WNQDVQELRAAT 335
QY 366 AGDA--VENFFNQTQNAIRIELRHQPIGRKLSGWGVQYLGQKSSALSATSSEAVKQPMLLD 423
Db 336 LGDARTVDMVFGLYRQNTN-----EKLNSAYDMPYLSSTGYTTAETLAA----- 382
QY 424 NKVQHYSPFGEVQANW-----DNFTLGGVVEKQASIRYDKALIDRENKYKQPLDLGAH 480
Db 383 -----YS-----DLTWHLTDREDDIGGVRFSDKSSSTQYHGSMLG-----NPFQDQGS 426
QY 481 RQTARFALSNGWYFTPQHKLSITASHQERLPSTQELYAHGKHVATNTFEVGNKHLNKR 540
Db 427 NDDQVLGQLSAGVMLTDDWRV-YTRVAQYKFSGNYI-----VPTAGLDARFVAEK 477
QY 541 SNNIELALYEGDRQWYNALYRNFEGYIYAQTINDGRGPKSIEDDEMKLVRYNSGA 600
Db 478 SINEYLGTRYE-----TADVTLOAATF--YTHTKDQMLYSGGPRVQMOTLS-----NAGKA 524
QY 601 DFVGAEGEYFPTPRYRIGVSGDYVRGRKLNPLSGREDAYGNRPFIADQDNAPRVP 660
Db 525 DATGVELEAKWFAFGWSWDINGVIRSEFTN-----DSELYHGNR-----VFPVP 570
QY 661 AARLGEHLKASITDRIDANLOYRVPQAUKLARYETRTFGHHMLMLGANYRNRYRGE-- 718
Db 571 RYAGSSVNGVIDTRYGA-----LMPFLAVNLVGHYFD--GDNQLRQGTATLTD 618
QY 719 ----W-----NWKYKADNLLNQ-----SVYAHSSFLSDTPQMGSRFTGGVNVK 757
Db 619 SSLGWAQATERMISVYVDNLPDRRYETGYMGNGSSAVAQVNMGRIVGINTR 669

RESULT 5
FYUA_YEREN STANDARD; PRT; 673 AA.
AC P46360;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticin receptor precursor (IRPC) (IPR65).
GN FYUA.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51871 / WA-314 / Serotype O:8;
RX MEDLINE=95075311; PubMed=7984105;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence
RL factor with dual function."
RL Mol. Microbiol. 13:253-263(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1223-75-1, 8081 / Serotype O:8, YE737, and 1209-79;
RA Rakin A., Heesemann J.;

```

```

RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
CC SIDEROPHORE YERSINIACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z29675; CAA82773.1; -
CC EMBL; Z35485; CAA84620.1; -
CC EMBL; Z35486; CAA84621.1; -
CC EMBL; Z35487; CAA84622.1; -
CC EMBL; Z35496; CAA84630.1; -
CC PIR; S60142; S60142.
CC HSP; P05825; 1PEP.
CC InterPro: IPR000531; TonB boxC.
CC Pfam; PF00593; TonB_dep_rec; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
CC Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
CC Transport; TonB box.
CC SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 673 PESTICIN RECEPTOR.
CC SITE 30 37 TONB BOX.
CC SITE 657 673 TONB C-TERMINAL BOX.
CC FT VARIANT 362 362 N -> D (IN STRAINS 8081, YE737 AND 1209-
CC FT 79).
CC SQ SEQUENCE 673 AA; 73682 MW; 1C20E0352FAD4DCC CRC64;

Query Match 5.2%; Score 210; DB 1; Length 673;
Best Local Similarity 21.7%; Pred. No. 6.2e-07;
Matches 180; Conservative 95; Mismatches 320; Indels 236; Gaps 43;

QY 1 MAQTLLKPIVLISILLINTPLLAQAHETEQSVGLTFTVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTRLPLALGGLLI--PAIANAQTQQDESTLEVT-ASKQSSRSAS-----ANNVSSTV 53
QY 61 ISGDTLRQKAVNLGALDGV-PGIHASQVGGGASAPV-IRGOTGRRIKVLNHHGTGDMA 118
Db 54 VSAPELSDAGVTASDKLPVLEGLNIENSGNMLFTSILRGVSSAQ-----99
QY 119 DF-SPDHALIWD-----TALS--QQVEILRGPTVLLYSSGNVAGLVADVADKTIPE 165
Db 100 DFNPAVTLVYDGVGPOLSTNTIQAULTDQSVELLRGPGQTLGKSAQGGIINIVTQO-PD 158
QY 166 KMP-----ENGVSCELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRK--SGDYAVPR 219
Db 159 STPRGVIEGGVSSRDSYR-SKFNL-----SGFIQDGL-----LYGSVTLRLQVDDGDMINPA 209
QY 220 YRN-----LKRLPD-----SPRRFANGQHRAVLGWRKRFYRTYSDR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPMWGMFAASRECTRATQDAYVGNW-----258
QY 257 RDQYGLPAHSHEYDDCHADIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDD 316
Db 259 -----DIKGRKLSISDGSPPDPMRRCT-----DSQTLSGKYTTDD- 293
QY 317 AHAHAHNGKFWIDLRNKRKYELAEWKQ-----PPGFPAALRVHLNRNDYHHDEK----- 365
Db 294 -----WV-----FNLSAWQOQHYSTFPF9--GSLIVNMPQR--WNQDVQELRAAT 335
QY 366 AGDA--VENFFNQTQNAIRIELRHQPIGRKLSGWGVQYLGQKSSALSATSSEAVKQPMLLD 423
Db 336 LGDARTVDMVFGLYRQNTN-----EKLNSAYNMPYLSSTGYTTAETLAA----- 382
QY 424 NKVQHYSPFGEVQANW-----DNFTLGGVVEKQASIRYDKALIDRENKYKQPLDLGAH 480

```

Db 383 -----YS-----DLTWHLTDREIDGGVRFSDKSSQYHGSMLG-----NPFQDQKS 426
 Qy 481 RQTARSFALSGNWYFTPOHKLSTASHOERLPSTQELVAHGKHVATNTFEVGNKHLNKR 540
 Db 427 NDDQVUGLSAGYMLTDHVR-YTRIAQYKESYNI-----VPTAGLDARFFVAEK 477
 Qy 541 SNNIELALGVGDRWQYNALVRNPGNVIYAOTLNDGRGPKSIEDSEMKLVRNOSGA 600
 Db 478 SINVELGTRYE-----TADVTLOAATF--YTHTKDMQLYSGPVGMQTLN-----NAGKA 524
 Qy 601 DFYGAEGEIVFKPTPRYRIGVSDYVRGLKRLPSLPGREDAYGNRPFIADQDQNAVRP 660
 Db 525 DATGVLEAKWRPAPGWDINGNVRSEFTN-----DSELYHGNR-----VPFVP 570
 Qy 661 AARLGHKLKASLTDRIANDLYVRFAQNKLYARTETPCHHMLNLGANYRNRTRYGE-- 718
 Db 571 RYAGAGSVNGVIDTRYGA-----LMPRLAVNLVGPVYFD-GDNQLRQGTATLND 618
 Qy 719 -----W-----NWYVKADNLNQ--SVYAHSSFLSDTPQMGSRFTGGVNVK 757
 Db 619 SSLQWQATERINISVHDNLDFERYTYGYMNGSSAVAQVNNGRVGINTR 669

RESULT 6
 HMUR YERPE STANDARD; PRT; 676 AA.
 ID Q56989;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemin receptor precursor.
 GN HMUR OR YPO0283 OR Y0543.
 OS Versinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Versinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM6;
 RX MEDLINE=96386041; PubMed=9026634;
 RA Horning J.M., Jones H.A., Perry R.D.;
 RA "The hmu locus of Versinia pestis is essential for utilization of
 RT free haemin and haem-protein complexes as iron sources.";
 RL Mol. Microbiol. 20:725-739(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
 RA Leather S., Moutie S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RA "Genome sequence of Versinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=42137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Ferhstern J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RA "Genome sequence of Versinia pestis KIM.";
 RL J. Bacteriol. 194:4601-4611(2002).
 CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
 CC UPTAKE BY BINDING HEMIN, AN IRON CHELATING SIDEROPHORE THAT ALLOWS
 CC THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U60647; AAC64866.1; --
 DR EMBL; AJ414141; CAC89146.1; --
 DR EMBL; AB013655; AAM84131.1; --
 DR PIR; AG0035; AG0035.
 DR PIR; T12069; T12069.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
 KW Complete proteome.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 676 HEMIN RECEPTOR.
 FT SITE 44 51 TONB BOX.
 FT SITE 659 676 TONB C-TERMINAL BOX.
 SQ SEQUENCE 676 AA; 74230 MW; 84ED731CB914ACD3 CRC64;

Query Match 5.1%; Score 205.5; DB 1; Length 676;
 Best Local Similarity 19.3%; Pred. No. 1.3e-06;
 Matches 164; Conservative 110; Mismatches 282; Indels 293; Gaps 37;

Qy 11 LSTLLINT-PLLAQAHEHQSVGLTETVVGVKSRPRATSGLLHTSTASDK----- 59
 Db 14 LSLAIACTLPLATQA-----ADTTTQTSSKHSHTDVTWVTATGNERSFEAPMMVT 65
 Qy 60 IISGDT-LRQKAVNLGDALDVGPIHASYGGGASAFVIRGOTGRRIKVNLNHHGTGDM- 117
 Db 66 VIEGNAPTQATAADMLRQVPLTVG-----SGTNGQDVVMRGYKQGVLT 115
 Qy 118 -----ADFSPDHAIMVDTALSOQVEILRPVTLTYSSGNVAGL-----VDVADGKIP 164
 Db 116 LVDGVRRQGTDTGHLNSTFLDPAIVKRIEIVRGPAALLYGSGALGVIAVETVDAAD---- 171
 C 165 EKPENGVSCELGLRLSSGNLEKLTSGGINIGKGNFVLHTE-----GLYRK 211
 Db 172 --MLQPGQNS--GYRVIS-----SAATGSHSFLGASAFGTDDGLGILSFGTRDIGNRQ 223
 Qy 212 SGDYAVPR-----YRNLRPLDSPRRFANGQHRVILGW 244
 Db 224 SNGFNAPNDETISNVLAKGTWQIDSIQSLSANLRYNNNSAIEPKNPOTSAPSSTN-VMTN 282
 Qy 245 RKRFRYRTYSDRDQYGLPAHSHEYDDCHADIIWOKSLINKRYLQLYPHLLTEEDVDYDN 304
 Db 283 RSTIQ-----DAQRYNFKPLQEWLNATAQVYSEVEINAR-----PGSABEGREQT 333
 Qy 305 PGLSCGFHDDDDAHAHNGKPMIDLRNKRYELRAEWKOPFPFGFALRVHL---NRNDYH 361
 Db 334 EGVK-----LENRTLF-----IESPASHLLTYGTETK 362
 Qy 362 HDEKAGDAVENF-----FNNQTONARIELHQPIRLKSGWGVQLGOKSSALSATSEA 415
 Db 363 QEQTPGATESFPQAKIRFSSGMLQDEITRLDPLVSILAGT----- 403
 Qy 416 VKQFMLLDNKVQHSYFQVQEQANWDNFTLEGGVRVVEKQKASIRYDKALIDRENYTKQPLP 475
 Db 404 -----RYDYSSSSGDY-----A 416
 Qy 476 DLGHRQCTARSFALSGNWYFTPOHKLSTASHQE--RLPSTQELYANGKHVA----- 525
 Db 417 DVDADKWSRG-AIS-----ITPTDMLMFGSAQAFAFTMGEMYNDSKHFPIRPGLT 471
 Qy 526 -TNTFEVGNKHLNKRNNIELALGYE-----GDRWOYNLALVRNPGNVIYAOTLND 577
 Db 472 LTN-YWVPENLKPETNETQCYGFGRLPFDLLMAEDDLQFKVSYFDTRAKDYISTEV--- 527

QY 578 GRGPKSIEDSEMMLVRYNOSGADFYGAEGEYIFKTPRYRIGVSGDYVGRGLKNLESLP 637
 Db 528 -----DQAMTTSVNDQAKINGWDSMSYK-TALFNWDLAYNTRGKNQ----- 573
 QY 638 GREDAYGNRPFFIAODQONAPR-----VPAARLGFHLKASLTDRIIDANLDYRVFAQNK 691
 Db 574 --TDEW-----LDTINPDVTSTVDVPVANGSF-----SVGMIGTEFA-NRS 611
 QY 692 ARYETETP--GHEMLNLGANYRNTRYGWNWVVKADNLNLOSVAHSSFLSDTPQMGRS 749
 Db 612 SRVSSSTPQAGYGVNDYFVSYKQOAFKGMTTMLLGNVFEKEYITPQGI-----PDGR- 666
 QY 750 FTGGVNVKF 758
 Db 667 -----NVKF 670

RESULT 7

HXC2_HAEIN STANDARD; PRT; 725 AA.
 AC P45357;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Heme/hemopexin utilization protein C precursor.
 GN HKUC.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D42 / Serotype B;
 RX MEDLINE=95270579; PubMed=7751272;
 RA Cope L.D., Yegor V.R., Mueller-Eberhard U., Hansen E.J.;
 RT "A gene cluster involved in the utilization of both free heme and
 heme/hemopexin by Haemophilus influenzae type b.";
 RL J. Bacteriol. 177:2644-2653 (1995).
 CC -!- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
 CONCENTRATIONS
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -!- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE D OF
 H. INFLUENZAE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U09840; AAA87059.1; -.
 DR PIR; A57148; A57148.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Transport; TonB Box; Signal; Receptor.
 FT SIGNAL 1 19 OR 21 (POTENTIAL).
 FT CHAIN 20 725 HEME/HEMOPEXIN UTILIZATION PROTEIN C.
 SQ SEQUENCE 725 AA; 80838 MW; FC7886E020CB5BCE CRC64;

Query Match 5.0%; Score 203; DB 1; Length 725;
 Best Local Similarity 19.8%; Pred. No. 2.1e-06;
 Matches 158; Conservative 134; Mismatches 300; Indels 206; Gaps 43;
 QY 11 LSTILNTPLLAQAHETEQSVGLTFTVVGKSRPRATSGLLHTSTASDKIISGDTLRQKA 70
 Db 6 LSLAIATT--LVTNALAQSVELDSINVIATRDP---SRFATPEKQSK---DSSLRSQA 57

RESULT 8

CIRA_ECOLI STANDARD; PRT; 663 AA.
 ID CIRA_ECOLI
 AC P17315;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Colicin I receptor precursor.
 GN CIRA OR CIR OR FEUA OR B2155.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89123100; PubMed=2644220;
 RA Nau C.D., Konisky J.;
 RT "Evolutionary relationship between the TonB-dependent outer membrane
 transport proteins: nucleotide and amino acid sequences of the
 Escherichia coli colicin I receptor gene.";
 RT

QY 71 VNLGDALDGVPGIHASQYGGG---ASAPVIRGOTGR-IKVLNHHGETGMADFPD-- 124
 Db 58 TSAAALEDIPNDIR--GGSRSIAQPNIRGSDNRVQVQIDGVRQNFDA-----HRG 110
 QY 125 AIVVDTALSQQVILRGPVTLTYSSNGVAGLVADGKIPEKMPENGVSGBELGRLSSG- 183
 Db 111 SYFLPMSLTOEIEVIGKPPSSSLGSGALGGVAMRTNALDLLKN--DKFGVKIRQGY 167
 QY 184 -NLEKLTSGGINTI-GLGKNFVLHTEGLYKSGDYAVPRYRNLRKLRLPSPRRFANGQHRV 241
 Db 168 QTANLISERDVSFAAMDKFDVLISGFYNNADNLRTGKGNKLAN--TAYKQFGG--LAK 222
 QY 242 LGWRKFYRT-YSDRRDQYGLPAHSHEYDDCHADIIWQSLINKRYLYQLYPHLLTEEDV 300
 Db 223 FGWQINDANRVELSHRETRFKQTAPS-----NEVENLTNEQITDOIR 266
 QY 301 DYDNPGLSCGFHDDDDAHAAHAKPMWIDLNRKRYELRAEWKQPPGFPEALRVHLNRNDY 360
 Db 267 EFHKP-----NNGSP-----PKAKPSQEEF-----YSGVKTFRGVS 299
 QY 361 HHDEKAGD--AVENFF--NNQTONARIELRHQPIGRKSGWGVQYLGQKSSALSATSEA 415
 Db 300 LTDOQIPDQSTVFNYLTPDNVYLNTHIALYNNKT-----IEKQKRVSGVKDQTKL 351
 QY 416 VKQPMLLDN--KVQHSYF-EGVEQANWDNFTLEGV-----RVEKQKA----- 455
 Db 352 TTRGINLRNSELSHISFVYGVGYMR-DKIRTERGTNNKDAQFRADPNANSNTTGVYLI 410
 QY 456 -----SIRYD-----KALIDRENTYKQPLDLAGHARTARSFALSGNWF 495
 Db 411 AHIFLGEKLLLSVSVYDHTVTSKTVYKDNHL-----SPATKLTWIV 455
 QY 496 TPQHKLSLTASHOB--RLPSTQELYAHGKHVAT-----NTFEVGNKHLNKRSNIE 545
 Db 456 T--NWLDFATKYNEAFRAPSQMERFVSGSHFGTSLGRNEINKF-VANPNLRPETAKKE 512
 QY 546 L-----ALGYEGDRWQYNLALYENRFGNVIYAQTLDNGRGSIEDDEMKLVRYNQS 598
 Db 513 ITANLHFDLSLFKQGDKFKIEATYFRNDVKDPINLKIFNDAKTNTNASASA-----GA 564
 QY 599 GADFVGAEGEYIFKPT-PRYR-----IGVSGDYVRGRL---KNLPSLPGRDAYCN 645
 Db 565 GA---GANPGALLPTKSOYQINWARSGLIELAQYQTERLTFTNYGSGTKGDKDSG- 620
 QY 646 RPIAODDQNAAPVPAARLGFHLKASLTD--RIDANLDYRVVFAQNKLYETET-POH 701
 Db 621 -----EALSNTAAASKIGVGVNYALVKDKFTVCATVTHYA--AQRVVKDHSVITYPSY 670
 QY 702 HMLNLGANYRNTRYGEW 719
 Db 671 ILTDLRATYA--PLKGEW 686

J. Bacteriol. 171:1041-1047 (1989).
 [2]
 RP REVISIONS.
 RA Nau C.D., Konisky J.;
 RA J. Bacteriol. 171:4530-4530 (1989).
 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / BHB2600;
 RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
 RA Church G.M.;
 RA Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA (Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474 (1997).
 [5]
 RN SEQUENCE OF 1-59 FROM N.A., AND SEQUENCE OF 26-46.
 RP MEDLINE=88058737; PubMed=3316180;
 RX Griggs D.W., Tharp B.B., Konisky J.;
 RA "Cloning and promoter identification of the iron-regulated cir gene
 of Escherichia coli.";
 RL J. Bacteriol. 169:5343-5352 (1987).
 [6]
 RN SEQUENCE OF 1-125 FROM N.A.
 RP STRAIN=BF2;
 RX MEDLINE=92250419; PubMed=1315732;
 RA Steffes C., Ellis J., Wu J., Rosen B.P.;
 RT "The lvsP gene encodes the lysine-specific permease.";
 RL J. Bacteriol. 174:3242-3249 (1992).
 [7]
 RN SEQUENCE OF 1-21 FROM N.A.
 RP MEDLINE=90264362; PubMed=2160948;
 RX Griggs D.W., Kaika K., Nau C.D., Konisky J.;
 RA "Activation of expression of the Escherichia coli cir gene by an
 iron-independent regulatory mechanism involving cyclic AMP-cyclic
 AMP receptor protein complex.";
 RL J. Bacteriol. 172:3529-3533 (1990).
 CC -!- FUNCTION: NOT YET KNOWN. POSTULATED TO PARTICIPATE IN IRON
 CC TRANSPORT. OUTER MEMBRANE RECEPTOR FOR COLICINS IA AND IB.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- INDUCTION: BY IRON AND BY A CYCLIC AMP/CYCLIC AMP RECEPTOR
 CC PROTEIN COMPLEX.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J04229; AAA75183.1; -;
 DR EMBL; U00007; AAA60531.1; -;
 DR EMBL; AE000304; AAC75216.1; -;
 DR EMBL; M19295; AAA23581.1; -;
 DR EMBL; M89774; AAA17054.1; -;
 DR PIR; B64984; QRECIQ.
 DR HSSP; P05825; 1PEP.
 DR ECO2DBASE; B068.1; 6TH EDITION.
 DR EcoGene; EG10155; cirA.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB DEPENDENT REC 1; 1.
 DR PROSITE; PS01156; TONB DEPENDENT REC 2; 1.
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
 KW Complete proteome.

FT	SIGNAL	1	25	
FT	CHAIN	26	663	COLICIN I RECEPTOR.
FT	SITE	31	38	TONB BOX.
FT	SITE	646	663	TONB C-TERMINAL BOX.
FT	CONFLICT	97	97	S -> D (IN REF. 6).
FT	CONFLICT	528	528	I -> N (IN REF. 1).
SQ	SEQUENCE	663	AA; 73895	MW; 2C68A45D4B5EB414 CRC64;

Query Match 4.6%; Score 187.5; DB 1; Length 663;
 Best Local Similarity 20.0%; Pred. No. 2.2e-05;
 Matches 172; Conservative 110; Mismatches 271; Indels 307; Gaps 42;

QY	6	LKPIV-----LSILINTPLLAQAHEHTEQSGVLETVTVVVKSRPRATSGLLHTSTASDKI	60
DB	4	LNPFRVRLCLSAISCAWPLAVDDG-----ETWVVTASS-----VEQNLDKAPASISV	53
QY	61	ISGDTLRQKAV-NLGDALDGVGIIHASQYGGGASAPVIRG-----QTGRRIK----	106
DB	54	ITQEDLQRKPVQNLKDLKEVGLTNEGDNKRGVSRINGLSSSYTLIILVDGKRVRNRA	113
QY	107	VLNHEGETGDMADFSFDHAIMVDYDTALSOQVEILRGFVTLTLYSSGNVAGLVVDVADGKIPEK	166
DB	114	VFRHN-----DFDLNW-IPVDSI--ERIEVVRGPMSSLYGSDALGGVVNIITKKIGQK	163
QY	167	MPENGVSGLGL-----RLSSGNLEKLTSGGINIGLGNFVLTHTGLYRKSGDYAV	217
DB	164	W-----SGTVTVDTTIIQEHDRDGTNGQFTSGPLIDGV-----LGM	201
QY	218	PRYRL-KELPDSRRFA--NGQHRVILGWRKFRYRTYSRRDOYG-LPAHSHEY---	269
DB	202	KAYGLAKREKDDPQNSTTTDTGETPRIEGFSR-----DONVEFATPNQNHDTAG	254
QY	270	-----DDCHADIIWQSLINKRY-----LQLYPHLLTEEDVDYDNPGLSCG	310
DB	255	YGFDRQDRDSDLDKNRLERQNYSVSHNGWDYGTSELXY-----GEKVENKNPCNSSP	309
QY	311	FHDDDDAHAAHANGKFWIDLR--NRRYELRAEWKQFPQCFEALRVHLNRNDVHHDEKAGD	368
DB	310	I-----TSESTVDGKYTLFLTAINGPLTVGGWNR-----HD-KLSD	345
QY	369	AVENFNNTQNAIRTELHQPIRLKSGWGVYLGQKSSALSATSEAVKQPMLLDNKVQH	428
DB	346	AV-----NLT-----GTSSTKTSAS-----QY	362
QY	429	YSFFGVEQANWDNFTLEGGVRVEKOKASIRYDKALIDRENYKQPLDGLGAHRQTARSFA	488
DB	363	ALFVDEDEWEIFEPLALTTCVRMD-----DHETYGHWSPRAYLVYNATDVT	409
QY	489	LSGNWYFTPOHKLISLTASHQERLPSTQELYAHGKHVATNTFE-----VGNKHLNKRNN	543
DB	410	VRGGW-----ATAF--KAPSLQL--SPDWTSSCRGACKIVGSPDLKPETSES	454
QY	544	IELALGYEDR-W-----QYNLALYRN-----RFGNYIYAOTLNDGRG	580
DB	455	WELGLYMGEEGKLEGVESVTVFRNDVKDRISISRTSDVNAAPGQNFVGFGTGANGRR	514
QY	581	PKSIEDDESMKLVRYNQSGADFYGABGEIYFPTPRYRIGVSGDYVGRKKNLPSLPGRE	640
DB	515	IPVFS-----YNNVNKARIQGVETELKIPNDSEKLSINYTYNDR	555
QY	641	DAYGNRPFTAQDDONAPRVPAPARLPHLKASLTDRIDANLDYYRVF-----AQNKIARY	694
DB	556	-----DVSNGENKPLSDLPHTTANGTLDWKPLALEDWVSFVSGHYTGQKRADSA	604
QY	695	ETRTPGHMLNLGANYRRNTRYGEWN-----WYVKAD-----NL-----LNSVVAHSS	738
DB	605	TAKTFCGYTI-----WNTGAQWQTVKDKLRAGVNLNLDKDLSDSDSYN-	649
QY	739	FLSDTFPMGRSFTGGYNNKPF	758
DB	650	-----EDGRRYFMAVDYRF	663


```

RESULT 9
ID HEMR YEREN STANDARD; PRT; 687 AA.
AC P31459;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hemin receptor precursor.
GN HEMR.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RX MEDLINE=93049186; PubMed=1425573;
RA Stojiljkovic I., Hantke K.;
RT "Hemin uptake system of Yersinia enterocolitica: similarities with
RL other TonB-dependent systems in Gram-negative bacteria.";
RN EMBO J. 11:4359-4367(1992).
RN [2]
RP REVISIONS.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RA Stojiljkovic I.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X68147; CAA48250.1; -
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_dep_Rc; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
CC Outer membrane; Iron transport; Transport; TonB box; Signal;
CC Receptor.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 687 HEMIN RECEPTOR.
FT SITE 44 51 TONB BOX.
FT SITE 670 687 TONB C-TERMINAL BOX.
SQ SEQUENCE 687 AA; 75226 MW; 69092EPF7F5DC08A CRC64;

Query Match 4.6%; Score 186.5; DB 1; Length 687;
Best Local Similarity 19.2%; Pred. No. 2,8e-05;
Matches 164; Conservative 94; Mismatches 262; Indels 333; Gaps 39;

QY 8 PIVLSILINTPLLAQHETESVGLTETVTVVVKGRPRATSGLLHTSTASDK----- 59
Db 13 PLSLAI-ACTLISLAVQAADTS-----STQT--NSKKRIADTWVVTATGNERSFEAPMM 63
QY 60 --IISGDT-LRQKAVNLGDALGVPGIASHASQYGGASAPVIGOTGRRKIKVLNHHGETGD 116
Dd 64 VTVWEADTPTSETATSDMLRNIFGLTVT--GSGR-----VNGQD-----VTLRGYKQGV 113
QY 117 M-----ADFSPTHAIMVDTALSSQVEILRGVPVTLVSSGNVAGL-----VDVADGK 162
Dd 114 LTLVDIGRGTTGHLNSFLPALVKRVEIVRGSALLYSGALGGVISTVDAADLL 173
QY 163 IPEKMPENGVSSELGLRLSSGNLEKLTSGGINIGLGNKPNVLHTE-----GLY 209
Dd 174 LPQG-----NSGRVYVSA-----AATGDHSGFLGASAFGRDTDVVDGILSGTRDIGNI 221

```

```

QY 210 RKSGDYAVP-----RYRNKLKL-PDSPPRFANGQRAVL 242
Dd 222 RQSDGFNAFNDETISNVLAKGTWRIDQISLANRYNNNSALEPKNPQTSAASNTLMT 281
QY 243 GWRKEFYRTYSDRDOYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDY 302
Dd 282 D-RSTIQR-----DAQLKYNIFLDOEWLNATRAQVYSEVINAR-----PQTPPEGRKQ 331
QY 303 DNPGLSCGFHDDDDAHAHNGKPMWIDLNRKRYELRAEMKQPPPGFEALRVHL---NRND 359
Dd 332 TTKG-----GK--LENKRLFT-----DSPASHLLTYGTEA 360
QY 360 YHDEKAGDAVENF-----FNNQTONARIELRHOPIGELKSGWGVQYLGOKSSALSATS 413
Dd 361 YKQEQTSGATESFPQADIRFGSGWLQDEITLRLDLPVSLAGT----- 403
QY 414 EAVQPMLLDNKVQHSYFFGVEQANWNTLLEGGVRVEKQASIRYDKALIDRENYKQP 473
Dd 404 -----RYDNYRGSSEGY----- 415
QY 474 LPDLGAHRQTARSFALSGNWTPTQHKLSLTASHQE--RLPSTQELYAHGKHVATN--- 527
Dd 416 -ADVADAKWSSR-----GAVSVTPTDMLMLFGSYAQAFRAPTMGEMYNDSKHFMSNIMGN 469
QY 528 ---TFEYGVNKHINKERSNNIELALGYEGDRWQYNLALYRNRFGNYIYAQTINDGRGPKSI 584
Dd 470 TLTNYWVFNPLKPKETNETQYGFGL-----RFNDLMMMA----- 503
QY 585 EDDSMKLVRYNOSGADFYGABGEIVFKPTPRYRIGVSGDYVGRKLKPLSPGREDAYG 644
Dd 504 EDDLOFKASYFDTKAKDYIST-----GVTDWDFGFG-----PG----- 535
QY 645 NRPFTAQDDQONAPRVPAAALGFHLKASLTDRIDANLDYRVFPAQNKLAARYETRTPGHML 704
Dd 536 -----GLYCKNCST--YSTNIDRAKINGWDATMTYQIQ-----WP 568
QY 705 NLGANYRR-----NTRYGEMWVYKADNLLN--QSVYAHSSFL-----SD 742
Dd 569 NLGLAYNTRGKNQNTNEMLDITNPDTVTSTLDVFPVANSFGAVGWIGTFADRSSRVSSG 628
QY 743 TPQMGRSFTGGVN 755
Dd 629 TPQAGY-----GVN 637

RESULT 10
ID BTUB SALTY STANDARD; PRT; 614 AA.
AC P37409;
DT 01-OCT-1994 (Rel. 30, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vitamin B12 receptor precursor.
GN BTUB OR STM41130.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=93079446; PubMed=1448622;
RA Wei B.Y., Bradbeer C., Kadner R.J.;
RT "Conserved structural and regulatory regions in the Salmonella
RT typhimurium btub gene for the outer membrane vitamin B12 transport
RL protein.";
RN Res. Microbiol. 143:459-466(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

```

Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking I., Nhan M., Waterston R., Wilson R.K.;
 "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
 Nature 413:852-856(2001).
 CC -!- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR BACTERIOPHAGE BP23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M89481; AAA27031.1; -;
 DR EMBL; AB008893; AAL22968.1; -;
 DR StyGene; SGI0030; btub.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Cobalt transport; Transport; TonB box; Signal;
 KW Receptor; Complete proteome.
 FT SIGNAL 1 20 BY SIMILARITY.
 FT CHAIN 21 614 VITAMIN B12 RECEPTOR.
 FT SITE 26 33 TONB BOX.
 FT SITE 597 614 TONB C-TERMINAL BOX.
 FT CONFLICT 81 81 G -> R (IN REF. 1).
 FT CONFLICT 134 134 R -> P (IN REF. 1).
 FT CONFLICT 183 183 N -> I (IN REF. 1).
 FT CONFLICT 280 280 R -> S (IN REF. 1).
 FT SEQUENCE 614 AA; 68525 MW; 9F51F601A615FD62 CRC64;
 SQ
 Query Match 4.6%; Score 184.5; DB 1; Length 614;
 Best Local Similarity 21.5%; Pred. No. 3.2e-05;
 Matches 140; Conservative 83; Mismatches 235; Indels 193; Gaps 35;
 29 QSVGLTETVTVGKSRPRATSGLLHTSTASDKIISGTLFRQKAVNLGDALDVGPGIHASQY 88
 21 QDTSPTLVVTVANRQQPSASVAPVTI---VTRQDIERQWSTVNDVRLRPGVDIAGS 77
 89 GGA--SAPVIRGOTGRRIKV-----LNHGETGMDADFPDHAIMVDLTALSQQVEILR 140
 78 GGAGQNSSIFIRGTNSSHVLVLIDGVRNLAVSGS-ADLS-----QFPVSLVQRIEYIR 131
 141 GPVTLTLYSGNVAGLVADVADGKIPEKMPENGVSGLGLRLSSGNLEKLTSG-GIN----- 194
 132 GPRSAIYGSDAIGGVNI-----ITRDNFGT--ELTAGWGSNSYQNY 172
 195 -----IGLKNFVLHTEGLYRKGDYAVPYRNLKRLPDSFRFPANGQRAVLGWRKR 247
 173 DISTQQQLGENTRATLIGDYETKGFVDVAVKGGTGMAQDPDRDGLSKTLGAL----- 226
 248 FVRRYTSRRDOYGLPAHSHBYDDCHADIIWQKSLINKRYLQIYPHLLTEEDVDVNDPL 307
 227 --EHTFSRWSGF---VRGSGYDN-----RTDYD----- 250
 308 SCGFHDDDDAHAAHNGKFWIDLRNKRVELRAEWKQPPFPFALRVHL-----NRNDYHH 362
 251 -----AYSPGSLIDTR-KLYSQWDAGLHFG-ERISQLVSSVSHSKDNYN 297
 363 DKAG-----DAVENFNNQTONARIELRHQPIGRKSGWQYLGQKSSALTSATSE 414
 298 DPHYGRYDTSATLDEMKG-YNVQWTNS-VVVGHGNVGA-----GVDW--QKQTTTPTGTY 348
 415 AVKQPMLLDNKVQHYSFSGVEQANWNTFTLEGVRVEKQKASIRYDKALIDRENTYKPL 474
 349 V---PEGYDQRNTGVYLTGLQLG--DFTLEAARSDDNS----- 383

475 PDLGAHROTARSFALSGNMYFTPOHKLSTASHQRLPSTQBYLAHGKHVATNTEVGNK 534
 384 -QFGRH---GTWQTSAGMEFYGRFTASYGTSYKAPNLGLQYGY-----YGNP 428
 535 HLNKRSNNIELALGYE---GDRQYNLALYRNFNGVYIYAQTLDNGRPGKSIEDDSEM 590
 429 NLNPKSKQWEGA--FEGLTAGVSWR--ISGYRN-----DIND-----MIDYDDL 470
 591 KLVRYNQSGADFYGAEGEYFKPTPRYRIGVSGDYV--RGRLNKLPSPGPR 639
 471 QKY-INEGKARIKGLEATNFDTGTLTHT-VSYDVDARNAITDTP-LPRR 518
 RESULT 11
 HPUB NEIMC
 ID HPUB NEIMC STANDARD; PRT; 810 AA.
 AC P96949;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemoglobin-haptoglobin utilization protein B precursor.
 GN HPUB.
 OS Neisseria meningitidis (serogroup C).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=135720;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DNM2 / Serogroup C / Serotype 2a;
 RX MEDLINE=97206152; PubMed=9157245;
 RA Lewis L.A., Gray E., Wang Y.-P., Roe B.A., Dyer D.W.;
 RT "Molecular characterization of hpuAB, the haemoglobin-haptoglobin-
 utilization operon of Neisseria meningitidis.";
 RL Mol. Microbiol. 23:737-749(1997).
 CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
 CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE.
 CC -!- SUBCELLULAR LOCATION: Outer membrane (potential).
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U73112; AAC44893.2; -;
 CC InterPro; IPR000531; TonB_boxC.
 CC Pfam; PF00593; TonB_dep_Rec; 1.
 CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Transport; TonB box; Signal; Receptor.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 810 HEMOGLOBIN-HAPTOGLOBIN UTILIZATION
 FT SITE 793 810 TONB C-TERMINAL BOX.
 FT SEQUENCE 810 AA; 90626 MW; 0D08DD79DA9CB817 CRC64;
 Query Match 4.6%; Score 184.5; DB 1; Length 810;
 Best Local Similarity 20.9%; Pred. No. 4.8e-05;
 Matches 190; Conservative 114; Mismatches 346; Indels 257; Gaps 49;
 7 KPVLVILLINT-PLLAQAHETEQSVGLTETVTVGKSRPRATSGLLHTSTASDKIISGPT 65
 6 KPVLAAVAIAQAFPAADPAQSAQTLEITVTGTHK-----TQKLGEEKIRKT 56
 66 LRQKAVNLGDALDGV---PGIHASQY-GGASAPVIRGOTGRRIKV----- 107
 57 LDKLLVN--DEHDLVRYDPGIGSVVVEGGRAGSGFTIRGVDKRVAINVDGLAQAESRSSE 114
 108 -----LNHGETGMDADFPDHAIMVDLTALSQQVEILRGPVTLTLYSGNVAGLV 156

```

Db      115 AFQELFGAYGNFNAWNTSEPNFS-----EVVITKGADSLKSGSGALGAV 161
Qy      157 DV-----ADGKIPKWPEN-GVSGELGLRLSSGNLEKLTSGGINICL-----GK-- 199
Db      162 NYQTKSASDVSEDKPYHLGKGG-SVGKNSQKSFSSITAAGRLFGDLALLVYTRFRGKET 220
Qy      200 -----NFVLHTGLRKSGDYAVP-RY-----RNLRKLPDPSRRFA----- 234
Db      221 KNRSTEGNEIKNDGYVNPNTDTGPGSKYLTLYATGVARSQDPQEWNKSTLFKLGYNF 280
Qy      235 NGQHRVGLWKRKFRYRTYSRRDQYGLPAHSHVEYDDCHADIIWQKSLINKRYLQLPHL 294
Db      281 NDQNR--IGW--IFESRTR-----FTNLSN-----LWTGTTSAATGD-YRH- 320
Qy      295 LTBEDVDY-DNPGLSGCFHDDDDAHAAHNGKWPIDLNRKRYELRAE-----WKQFPFGFEA 350
Db      321 --RDQSVYRRSGVE--YKNELEHG-----PWDSLKLRYDKQRIWDMTWTWIDPIKNYD 369
Qy      351 LRVLHNRNDYHDEKAGDAVENFFNQTON-----ARIELRHQPIGRKSGNGVOY---L 402
Db      370 LR-GINSEVYHS-----FHRIRONTAOWTADFE---KQLDFSKAVMAAQYGLGG 414
Qy      403 GQKSSALS-----ATSEAVKQPMLLDNKVQHYFFGVGEQANWDFTEGGVR 449
Db      415 GRGNANSYDYFVKLYDPKILTSNOAKITMLIENR-SKYF-----AYWNVHFLGND 468
Qy      450 VEKQASIRYDK---ALIDRENY---YKQPLPDLGAHQRTAR-SFALSGNMYFTPOHKL 502
Db      469 RFRLNAGIRYDKNSSAKDDPKYTAIRGQIPHLGSERAHAGFSVGTGDFWRTKHLHL 528
Qy      503 LTASHQERLSTQELYAHGKHVAINTPEVGNKHINKERSNNIELALGVEGRWYNALY 562
Db      529 AKYSTGFRAPTSDETLLFPH--PDFYIKANPLNKAERKAWELAGSAGKAGNFKLSGF 586
Qy      563 RNREGNYI-----YAQTLNDGRGPKSTEDSEMKLRYNQS-----G 599
Db      587 KTKYRDFIELTYMGVSSDNPNKPYA-PLSDGTALVS-----SPVWQNRSSAWVK 638
Qy      600 ADFVGAGEIYFKPTPRYRGVSGDYVRGRUKNLPSPGREDAYGNRPFFIAQDDQNA 657
Db      639 LEFGTWNLSDIGLPQGTAGVNVSYIKGKAK-----QTNGQE-----TPINALSP 690
Qy      658 ---RVPARGLFHLKASIT-----DRIDANLYYRVFAQNKLARYETRTPGHMLNIG- 707
Db      691 LGYDAPSKRWGINAYARTAAKESDTHSNDLNNPWPYAKHASKAYTLFDLSAYLNIGK 750
Qy      708 -----ANYR-RNTRYGEW-----NMVVKADNLLNQSVYAHSSFLSDTPQWGRSFT 751
Db      751 QVTLRAAXYNTNKQYVTWESLRSIRBEFGTVNRVDNKTTHAGIQRTS-----PGRSYN 803
Qy      752 GGVNVKF 758
Db      804 FTIBAKF 810

```

RESULT 12

```

BTUB_ECOLI
ID BTUB_ECOLI STANDARD; PRT; 614 AA.
AC P06129;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vitamin B12 receptor precursor.
GN BTUB OR BFE OR CER OR DCRC OR B3966.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85130824; PubMed=3882670;
RA Heller K., Kadner R.J.;

```

"Nucleotide sequence of the gene for the vitamin B12 receptor protein in the outer membrane of *Escherichia coli*."; J. Bacteriol. 161:904-908(1985).

[2]

SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
MEDLINE=94089392; PubMed=8265357;
RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.;

"Analysis of the *Escherichia coli* genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes."; Nucleic Acids Res. 21:5408-5417(1993).

[3]

SEQUENCE OF 1-5 FROM N.A.
MEDLINE=91154132; PubMed=1999392;
RA Gustafsson C., Lindstroem P.H., Hagervall T.G., Esberg B., Bjork G.R.;

"The trmA promoter has regulatory features and sequence elements in common with the rRNA P1 promoter family of *Escherichia coli*."; J. Bacteriol. 173:1757-1764(1991).

[4]

SEQUENCE OF 456-614 FROM N.A.
STRAIN=RDD020;
MEDLINE=93106943; PubMed=8093236;
RA Dougherty T.J., Thanassi J.A., Pucci M.J.;

"The *Escherichia coli* mutant requiring D-glutamic acid is the result of mutations in two distinct genetic loci."; J. Bacteriol. 175:111-116(1993).

[5]

SEQUENCE OF 21-32.
STRAIN=K12 / EMG2;
MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;

"Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12."; Electrophoresis 18:1259-1313(1997).

[6]

MUTAGENESIS OF TONB BOX.
MEDLINE=90078094; PubMed=2687240;
RA Gudmundsdottir A., Bell P.E., Lundrigan M.D., Bradbeer C., Kadner R.J.;

"Point mutations in a conserved region (TonB box) of *Escherichia coli* outer membrane protein Btub affect vitamin B12 transport."; J. Bacteriol. 171:6526-6533(1989).

!- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR BACTERIOPHAGE BP23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.

CC

!- SUBCELLULAR LOCATION: Outer membrane.

CC

!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

EMBL; M10112; AAA23524.1; --
EMBL; U00006; AAC43072.1; --

DR

EMBL; AE000471; AAC76948.1; --
EMBL; M57568; -- NOT ANNOTATED_CDS.

DR

EMBL; L14556; AAA23676.1; --
PIR; A65204; QRECBT.

DR

ProGene; EG10126; btuB.
InterPro; IPR000531; TonB_boxC.

DR

Pfam; PF00593; TonB_dep_rec; 1.
PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.

DR

Outer membrane; Cobalt transport; Transport; TonB box; Signal;
Phage recognition; Receptor; Complete proteome.

KW

SIGNAL 1 20
FT CHAIN 21 614 VITAMIN B12 RECEPTOR.
FT SITE 26 33 TONB BOX.

```
FT SITE 597 614 TONB C-TERMINAL BOX.
FT MUTAGEN 28 30 L->P: INACTIVATE UPTAKE.
FT MUTAGEN 30 30 V->G: INACTIVATE UPTAKE.
FT CONFLICT 162 162 A -> G (IN REF. 1).
FT CONFLICT 377 377 A -> R (IN REF. 1).
SQ SEQUENCE 614 AA; 68407 MW; AB43CC46A991FF95 CRC64;

Query Match
Best Local Similarity 19.8%; Pred. No. 0.00013;
Matches 132; Conservative 78; Mismatches 211; Indels 246; Gaps 31;

Qy 29 QSVGLTETVVVK--SRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGALDGVFGIIAS 86
Db 21 QDTSPTLVVTVANRFEQPRST-VLAPTIV---VTRQDIDRWQSTSVNDVLRRLSGVDIT 75

Qy 87 QYGGGA--SAPVIRGOTGRRIKV-----LNHHGETGDMADSPDHAIWVDLTALSOQVEI 138
Db 76 QNGSGQLSSPIRGTNASHVLVDIGVRLNLAGVSGS-ADLS-----QPPIALVQRVEY 129

Qy 139 LRGPVTLTYSSGNVAGLVADGKIPEKMPENGVSGLRLSSGNLEKLTSGGINIGLG 198
Db 130 IRGPRSAVYGSDAIGGVNIIITR-----DEPQTEISA-----GWG 165

Qy 139 KNFVLHTEGLYKSGDYAVPRNL-----KRLPDSPPRFANGQRAVLGWRKRYRRTY 253
Db 166 SN-----SYQNDVSTQQQGDKTRVTLG DY----- 192

Qy 254 SDRRDQVGLPAHSHEYDDCHADIIWQSLINKRYLQLYPHLLTEEDVDYDNPGLSCGF-- 311
Db 193 -----ATHGYD-----VVAYGNTGTQAGTDND-----GFLS 219

Qy 312 -----HDDDA-----HAHANGKEPIDLR---NKRYELRAEWKOP 344
Db 220 KTLYGALNEHFTDAWSGFRVGYDNRITNDAYISPGSPLLDTRKLYSQSDAGLRYNGE 279

Qy 345 FGFEPALRVHLNRNDVHDEKAG-----DAVENFFNNTQNAIRIELRHOPIGELKGS 396
Db 280 LIKSQDITSYSHSKYNDYDPHYGRYDSSATLDEMKNQYTVOWANN---VIVHGSIGA--- 333

Qy 397 WGVQYILGQKSSALSATSEAVKQPMLLDNKVQHSYFEGVEQANW-----DNFTLRG 446
Db 334 -GVDM--QKQTTPTGTG-----YVEDGYDQRTGIVLTGLQVQGDFTFEG 375

Qy 447 GYRVEKQASIRYDKALIDRENYIKOPLDLAGHROTARSFALSGNWTTPHOKLSLTAS 506
Db 376 AARSDDNS-----QFGRH-----GTWQTSAGWEPFEGYRFTASYG 410

Qy 507 HQERLPSTOELYAHGKHVATNTFEVGNKHLNKNERSNNIELALGYE---GDRWQYNLALY 562
Db 411 TSKAPNLGQLYGF-----YGNPNLDPEKSKQWEGA--FEGLTAGVNMW--ISGY 456

Qy 563 RNRFGNYIIYAQTLDNGRGPKSEDDSEMKLVRYNQSGADFYGAEGEIIYFKPTPRYRIGVS 622
Db 457 RNDVSLI-----DYDDHTLKYNEGKARIKGVATANFDTGPLHT--VS 500

Qy 623 GDYVGR 629
Db 501 YDYVDAR 507
```

RESULT 13

```
YINCD_ECOLI
AC P76115;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable tonB-dependent receptor yncD precursor.
GN YINCD OR B1451.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
FA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.;
RA Riley M.; Collado-Vides J.; Glasner J.D.; Rode C.K.; Mayhew G.F.;
RA Gregor J.; Davis N.W.; Kirkpatrick H.A.; Goeden M.A.; Rose D.J.;
RA Mau B.; Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000242; AAC74533.1; -.
DR FIC; F64897; F64897.
DR EcoGene; EG13774; yncD.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB dep Rec; 1.
DR PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE NEG.
DR PROSITE; PS01156; TONB DEPENDENT REC 2; FALSE NEG.
KW Hypothetical protein; Outer membrane; Receptor; Signal; TonB box;
KW Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 700 PROBABLE TONB-DEPENDENT RECEPTOR YNCD.
FT SITE 680 700 TONB C-TERMINAL BOX.
SQ SEQUENCE 700 AA; 77260 MW; 7B3B96C6ABA48FE4 CRC64;

Query Match
Best Local Similarity 19.2%; Pred. No. 0.00019;
Matches 166; Conservative 122; Mismatches 292; Indels 283; Gaps 41;

Qy 5 TLKPIVLSILLINTPLLAQHETEQSVGLTETVVVKSPRATSGLLHTSTASDKLIISGD 64
Db 5 SVRQTVLPALLVSLPVPVFADEQ-----TMIVSAAPQVVSLE--DTPAAVSVDGE 53

Qy 65 TLR--QKAVNLGDALDGVPGI---HASQYV-----GGASAPVIRG---QYGRRI 105
Db 54 ENRLATPRNLSESLTVPGLQVQRNQAQDLQLSIRGSGRSTYIGIRLYVDGIPA 113

Qy 106 KVLNHHGETGDMADSPDHAIWVDLTALSOQVEILRGVPTLLYSSGNVAG---LVDVADGK 162
Db 114 TMDPGQGQTSN-----IDLSSVQNVVEVLRGPFPSALY--GNASGVNMVTTQTGQ 160

Qy 163 IPEKMPENGVSGL-----GLRLSSGNLEKLTSGGINIGLKNFVLTSLGYRKSQDYAV 217
Db 161 QPPTTEASSYSGPSGSRWYGLKATGATGDTGTFQGDV-----DYTV 200

Qy 218 PPRNLKRLPDSPPRFANGQRAVLGWRKRYRRTYSDRRDQVGLPAHSHEYDDCHADII 277
Db 201 -----STTRFT-----THGYRD-HSGA- 216

Qy 278 WOKSLINKRY-----LQLYPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHAKPWI 328
Db 217 -QKNLANAKLVGRIDEASKLSL---IFNSVDIKADDPG---GLTKAE-----WKANPQ 263

Qy 329 DLNRKRYELRAEWKOPFPFGFEALRVHLNPNNDVHDEKAGDAVENFNNTQNAIRIELRHQ 388
Db 264 APRASQYDTRTKTIKOTQAGLRYSRLSSKDDMSVMYAGERETTOYQSIPMAPQLNPSH- 322

Qy 389 PIGRLKSGMV-----QYLQKSSALSATSEAVKQPMLLDNKVQHSYFPGV----- 434
Db 323 -----AGGVITLQRHYOG-----IDSRWTHRGELGVPTFTTGLN 357

Qy 435 -----EQANWDNFTLEGGRVVEKOKASIRYDKALIDRENYIKOPLDLAGHROTARSFA 488
```

Db 358 YENMSNRKGYNFRNLNGMPEYQKGELRDE-----RNLMMNIDPYLQWQQLSEKLS 412
QY 489 L-----SGNYYFTPOHKL-----SLTASHOERLPSTQELY-----AHGKHA 525
Db 413 LDAGVRYSSVWFSDNDHYVTPGNGDSDGASHVKWLPAGSLKYAMTDANNIYLAAGRGE 472
QY 526 TNTFE-----VGNKHLNKNRSNNIELALGYE---GDRWOYNIALYRNFRGNYYIAQT 574
Db 473 TPTINELSYRADQSGMNLGLKLPSTNDTIEGSKTRIGD-GILLSALFQTDITDDEIVVDS 531
QY 575 LNDGRGPKSIEDDSEMLVRYNQSG-ADPYGAEIGIYFKPTPRYIGVSGDVVRGLKNL 633
Db 532 SSGGR-----TTYKNAGTRQGAELAWDQRFAGDFRNASWTWLDATYRS- 577
QY 634 PSLPGREDAYGNR-PFIAQDDQNAVPAPARLGFHLKASLTDRIDANLDYVRV-----FAQN 689
Db 578 -NVNCEQDCNGNRMPGIARN-----MGF-----ASIGYPEDGWYAGT 614
QY 690 KIARY-----ETPTPGHMLNLGANYRNTRYGEWNVYKADNLLNQSVYAHSSF 739
Db 615 E-ARYMGDIMADDENTAKAPSYTLVGLFTGYKYNYHNLTVDLFGRVDNLFDEK-YVGSVI 672
QY 740 LSDT-----PQMGSEFTGGVNV 756
Db 673 VNESNGRYEYSPGRNYGVGMNI 695

RESULT 14
HPUB_NEIMA
ID HPUB_NEIMA STANDARD; PRT; 810 AA.
AC Q9UWA2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin-haptoglobin utilization protein B precursor.
GN HPUB OR NMA0474.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491.";
RL Nature 404:502-506(2000).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL162753; CAB83769.1; --
DR PIR; A81965; A81965.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.

DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Signal; Receptor;
Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 810 HEMOGLOBIN-HAPTOGLOBIN UTILIZATION
PROTEIN B.
FT SITE 793 810 TONB C-TERMINAL BOX.
SQ SEQUENCE 810 AA; 90570 MW; D38DE1DCA3CASA6E CRC64;
Query Match 4.2%; Score 169.5; DB 1; Length 810;
Best Local Similarity 20.6%; Pred. No. 0.00053;
Matches 186; Conservative 122; Mismatches 345; Indels 251; Gaps 53;
QY 7 KPVLVLSILLINT-PULLAQAHETEQSVGLETTVVVGVKSRPRATSGLLHTSTASDKIISGT 65
Db 6 KPVLAALAAIAQAFPAADPAQSAQTLEITVTGTHK-----TQKLGEEKIRRT 56
QY 66 LRQKAVNLGDALDGV---PGIHASQYG-GGASAPVIRGQTGRRIKV----- 107
Db 57 LDKLIVN--DEHDLVRYDPGISVVEGRAGSGFTIRGVYDKDRVAINVDGLAQAESRSSE 114
QY 108 -----LNHHGETGDMADRFSPDHALIMVDLTALSOQVEILRGPVTLTYSSGNVAGLV 156
Db 115 AFQELFGAYGNFNAIRNTSEPFN-----EVTITKGADSLKSGSGALGGAV 161
QY 157 DV-----ADGKIPEKMPEN-GVSGELGRLSLSGNLEKLTSGGINIGLGNFVL----- 203
Db 162 NVQTKSASDYYVSEDKPYHLGIKGG-SVGKNSQKFSISITAAAGRLFGLDALLVYTRRGKET 220
QY 204 ---HTEG-----LYRKSGDYAVPRY-----RNLKRLDPSRPPA----- 234
Db 221 KNRSTEGDIEIKNDGYVYNPTDTGGPSKYLTYVATGVARSQDPDQEWNKSTFLKLYNF 280
QY 235 NGQRAVLGWRKRFVRRVTVSDRRDQVGLPAHSHVEDDCHADIIWQSLINKRYLQLYPHL 294
Db 281 NDQNR--IGW---IFEDSRDTR-----FNELSN-----LWTGTTTSAATGD-YRH- 320
QY 295 LTEDVDY-DNPGLSGCFHDDDDAHAAHNGKPMIDLRNKRYELRAEWKQPPFGEALRV 353
Db 321 --RQDVSYRRSSGVE--YKNELEHG-----PWDSLK-LRYD-----KQ-----RI 355
QY 354 HLNRNDY-----HDEKA--GDAVENF--FNNTQWARIELRHQPIGRLGKSGVQVY---L 402
Db 356 DMNTWTWDIPKNYDKRNGINGEVYHSFRHRIQNTAQWTDATDFEQ-LDFSKAVAAQYGLGG 414
QY 403 GOKSSALS-----ATSEAVKQPMLLDNKNVQHSYFIVGEQANWDN-FTLEGVY 448
Db 415 GKGDNANSYSYFALKYDPKILASNOAKITMLIENR-SKYKF-----ATWNAFHLGGND 468
QY 449 RVEKOKASIRYDK---ALIDRENY---YKQPLDILGAHRQTAR-SFALSNGNMYFTPHKL 501
Db 469 RF-RLNAGIRYDKNSSAKDDPKYTTAIRGQIPHLGSEAHAGFSVGTGDFNRFTKHLHL 527
QY 502 SLTASHOERLPSTQELYAHGKHVAINTFEVGNKHLNKNRSNNIELALGVEGDRWQVNLAL 561
Db 528 LAKYSTGFRAPTSDETLLFPFH--PDFYLKANPNLKAFAKAKNWLGLAGSGRAGNFKLGG 585
QY 562 YRNRFGNYI---YAQTLNDGRGPKSIEDDSEMLV-----RYNOSGADFYGAEGEIVF 611
Db 586 FTKYRDFIELTYMGVSDDDKKNPRYAPLSGDTALVSSPVWQNRSAWVGIE-----F 641
QY 612 KPT-----PR-YRIGVSGDYVRGLKNLPSLFGREDAYGNRPFIAQDDQNAF----- 657
Db 642 NGTNWLDISGLPKGLHTGLNVSYIKG--ATQNNKGK-----TPINALSPWTAVYSLGY 693
QY 658 RVPAAALGFHLKASLT-----DRIDANLDYVRYPQAKNLARYETRTGHEMLNIG----- 707
Db 694 DAPSKRWGINAYATTAAKPSDTVHSNDDLNNPWPYAKHSAKAYTLFDLSAYLNIQKVT 753
QY 708 ---ANYR-RNTRYGEW-----NWYVKADNLLNQSVYAHSSFLSTPQWGRSFTGV 754
Db 754 LRRAAYNITNKQYTWYTESLRSIREFGTVNRVDNKTAGIQRTS-----PGRSYNFTI 806

```

QY 755 NVKF 758
Db 807 EAKF 810

RESULT 15
RHTA RHIME STANDARD; PRT; 746 AA.
AC Q923Q5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rhizobium receptor precursor (TonB-dependent siderophore receptor
DE rhtA).
GN RHTA OR RAI265 OR SMA2414.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=21172875; PubMed=11274118;
RA Lynch D., O'Brien J., Welch T., Clarke P., Cuiv P.O., Crosa J.H.,
RA O'Connell M.;
RT "Genetic organization of the region encoding regulation, biosynthesis,
RT and transport of rhizobactin 1021, a siderophore produced by
RT Sinorhizobium meliloti."
RL J. Bacteriol. 183:2576-2585(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -!- FUNCTION: RECEPTOR FOR THE SIDEROPHORE RHIZOBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF110737; AAD09419.1; -
CC DR EMBL; AE007312; AAK65923.1; -
CC DR PIR; A95420; A95420.
CC DR PIR; T46821; T46821.
CC DR InterPro; IPR000531; TonB boxC.
CC DR Pfam; PF00593; TonB_dep_Rec; 1.
CC DR PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE NEG.
CC DR PROSITE; PS01156; TONB DEPENDENT REC 2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Plasmid; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 746 RHIZOBACTIN RECEPTOR.
FT SITE 40 47 TONB BOX.
FT SITE 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80633 MW; 16AB44A4025D5B23 CRC64;

Query Match 4.0%; Score 161; DB 1; Length 746;
Best Local Similarity 18.9%; Pred. No. 0.0018;
Matches 165; Conservative 124; Mismatches 306; Indels 278; Gaps 41;

```

RESULT 16

FEPA ECOLI

ID FEPA ECOLI

AC P05825; P76821; P77093;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ferrienterobactin receptor precursor (Enterobactin outer-membrane

DE receptor).

GN FEPA OR FEUB OR B0584.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Escherichia.
RN [1]
RP NCBI_TaxID=562;
RX SEQUENCE FROM N.A.
RA MEDLINE=86278160; PubMed=3015941;
RA Lundrigan M.D., Kadner R.J.;
RT "Nucleotide sequence of the gene for the ferrienterochelin receptor
RT PepA in Escherichia coli. Homology among outer membrane receptors
RT that interact with TonB.";
RL J. Biol. Chem. 261:10797-10801(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
RA Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 1-77 FROM N.A.
RX MEDLINE=89066678; PubMed=2974033;
RA Pettis G.S., Brickman T.J., McIntosh M.A.;
RT "Transcriptional mapping and nucleotide sequence of the Escherichia
RT coli fepA-res enterobactin region. Identification of a unique
RT iron-regulated bidirectional promoter.";
RL J. Biol. Chem. 263:18857-18863(1988).
RN [6]
RP MOLECULAR ANALYSIS.
RX MEDLINE=90354449; PubMed=2201687;
RA Armstrong S.K., Francis C.L., McIntosh M.A.;
RT "Molecular analysis of the Escherichia coli ferric enterobactin
RT receptor PepA.";
RL J. Biol. Chem. 265:14536-14543(1990).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=93101384; PubMed=9686293;
RA Buchanan S.K., Smith B.S., Venkatramani L., Xia D., Esser L.,
RA Palnitkar M., Chakraborty R., van der Helm D., Deisenhofer J.;
RT "Crystal structure of the outer membrane active transporter PepA from
RT Escherichia coli.";
RL Nat. Struct. Biol. 6:56-63(1999).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC UPTAKE BY BINDING FERRIENTEROACTIN (FE-ENT), AN IRON CHELATIN
CC SIDEROPHORE THAT ALLOWS E. COLI TO EXTRACT IRON FROM THE
CC ENVIRONMENT. PEPA ALSO ACTS AS A RECEPTOR FOR COLICINS B AND D.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M13748; AAA65994.1; -;
DR EMBL; A5000163; AAC73685.1; -;
DR EMBL; U82598; AAB40783.1; ALT_INIT.
DR EMBL; D90700; BAA35225.1; -;
DR EMBL; J04216; AAA23756.1; -;
DR PIR; F64791; QRECFP.
DR PDB; 1FEP; 13-JAN-99.
DR ECODBASE; D079.0; 6TH EDITION.
DR EcoGene; EG10293; fepA.
DR InterPro; IPR000531; TonB boxC.
DR Pfam; PF00593; TonB dep Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Transmembrane; Outer membrane; Iron transport; Transport; TonB box;
KW Signal; Receptor; 3D-structure; Complete proteome.
FT SIGNAL 1 22
FT CHAIN 23 746 FERRIENTEROACTIN RECEPTOR.
FT SITE 34 41 TONB BOX.
FT SITE 729 746 TONB C-TERMINAL BOX.
FT CONFLICT 152 152 A -> R (IN REF. 1).
FT CONFLICT 403 403 MISSING (IN REF. 1).
FT HELIX 40 44
FT TURN 45 46
FT TURN 48 49
FT STRAND 50 54
FT HELIX 55 60
FT TURN 67 67
FT HELIX 68 71
FT TURN 72 73
FT TURN 75 76
FT STRAND 77 81
FT TURN 84 85
FT TURN 87 90
FT STRAND 92 96
FT TURN 97 98
FT HELIX 101 103
FT STRAND 104 108
FT TURN 109 110
FT STRAND 111 112
FT HELIX 115 118
FT TURN 123 124
FT TURN 133 134
FT HELIX 138 140
FT STRAND 141 148
FT TURN 149 149
FT HELIX 150 153
FT TURN 154 155
FT TURN 157 158
FT STRAND 162 168
FT STRAND 176 186
FT TURN 189 190
FT STRAND 194 204
FT STRAND 209 220
FT TURN 225 231
FT STRAND 235 236
FT HELIX 237 239
FT TURN 240 241
FT STRAND 249 264
FT TURN 265 266
FT STRAND 267 281
FT HELIX 294 298
FT TURN 299 300
FT STRAND 304 318
FT TURN 320 321
FT STRAND 324 338
FT STRAND 360 378
FT STRAND 382 397
FT STRAND 426 441

```
FT STRAND 446 457
FT TURN 458 460
FT STRAND 461 473
FT TURN 476 477
FT STRAND 478 489
FT TURN 493 496
FT STRAND 498 499
FT TURN 501 503
FT STRAND 517 519
FT TURN 522 523
FT STRAND 527 540
FT TURN 541 542
FT STRAND 543 560
FT STRAND 567 567
FT TURN 569 570
FT STRAND 573 597
FT TURN 598 599
FT STRAND 600 614
FT TURN 615 617
FT STRAND 627 638
FT TURN 639 640
FT STRAND 641 650
FT STRAND 653 653
FT STRAND 658 658
FT TURN 660 661
FT STRAND 664 664
FT HELIX 667 670
FT STRAND 673 673
FT STRAND 676 686
FT STRAND 691 698
FT TURN 700 701
FT STRAND 705 705
FT STRAND 714 715
FT TURN 717 718
FT STRAND 722 723
FT TURN 727 728
FT STRAND 733 733
FT STRAND 737 746
FT SQ SEQUENCE 746 AA; 82107 MW; 09348AAB1C29710A CRC64;

Query Match
Best Local Similarity 4.0%; Score 160.5; DB 1; Length 746;
Matches 135; Conservative 87; Mismatches 239; Indels 257; Gaps 34;

QY 14 LLINTPL--LAQAEH-TEQSVGLTETVTVCKSPRATSGLLHTSTASDKLIISDITLQKA 70
Db 10 LLVNLGIYGAQAEPTDTPVSHDDTIVVTAEQNLQAPGVSTITA-----DEIRKNP 62
QY 71 V--NLGDALDGVPGIHASQYGGGASAPVIRGQTRRIKVLNHHGETGDMADFPDHAIMV 128
Db 63 VARDVSKIIRTWPGVNL--GNSTG-----GQGGNNRQI-----DIRGMPENTLIL 107
QY 129 -----DTA-----LSQVEILRGVPTLLYSNGVAGLVDA 159
Db 108 IDGKPVSSNRVQRGWRGERTRGDTSWVPPEMIERIEVLRGFAAARYGNGAGGVYNI 167
QY 160 DGK-----IPEKMPEN-----GVSGELGLRLS---SGNLEKLTSGGINI 195
Db 168 TKKGSGEWGSDWAYENAPHEKEGATKTNLSLTGLDGFSEFRILYGNLDTKTQADWDI 227
QY 196 GLGKPNVLTGELGKRSQYAVPRYRNKLRLPDSPPRFANGQHRVILGW-----RK 246
Db 228 NQG-----HOSA---RAGTVAT-----TLPAGREGVINKDINGVVRWDFAPLQSLLE 272
QY 247 REYRR-----TYSDRDQICLPAHSHYDDCHADII-----WQSLNKRVLQLY 291
Db 273 AGYSROGLNLYAGDTQNTNSDSYTRSKYGDETNRLYEQNYALTWNGGWDNGVTTSNNWQ-Y 331
QY 292 PHL-----LTEDVDYNPGLSCGFHDDDDAHAAHAKGKPIDLR 331
Db 332 EHTNRSRIPEGLAGGTGEGKNEKATQDFVDIDL-----DDVNLHSEVNLPIDFILV 381
QY 332 NKRYELRAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFFNQTQNRQNIELRHQPTG 391
```

```
Db 382 NOTLTILGTEWQQ-----ORMKDLSSN-----TQ----- 404
QY 392 RLKSGWGVQYLGQKGSALSATSSEAVKQPMMLDNKVQHYSGFFGVEQANWONFTLEGGVRVE 451
Db 405 ALTGT-----NTGCAIDGVSTDRSPY---SKAEIFSLFAENNMELTDSTI----- 447
QY 452 KOKASIRYDKALIDRENYKQPLDGLGAHROFARSFALSGNMYFTPOHKLSLTASHOERL 511
Db 448 -VTPGLRFDDHHSIVGNW-----SPALNISQGLGDDFTL-----KMGIARAY--KA 490
QY 512 PSTQE-----LYAHGKHV---ATNTFEVGNKHLNKRSENNIELALGYEGDRQYNLALY 562
Db 491 PSDYQTNPNYILYSKGGQGYASAGGCYLQGNDDLKAETSINKBIEGLFKRDKGLAGVTWF 550
QY 563 RNRFGNYIYAQTLNDGRGPKSIEDDSEMKLVRYNQSGADFY-----CAEGEI 609
Db 551 RNDYRNKIEAGYVAVGQ-----NAVGTDLYQWDNVPKAVVEGLESL 592

RESULT 17
TBPI_HAEMIN STANDARD; PRT; 912 AA.
AC P44970;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable transferrin-binding protein 1 precursor.
GN TBPA OR TBPI OR HI0994.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Pine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD."
RL Science 269:496-512 (1995).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, U32780; AAC22656.1; -.
CC FIR, C64107; C64107.
CC TIGR, HI0994; -.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_dep.Rec. 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC.1; 1.
CC PROSITE; PS01156; TONB_DEPENDENT_REC.2; 1.
CC Outer membrane; Receptor; Signal; TonB box; Complete proteome.
CC SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 912 PROBABLE TRANSFERRIN-BINDING PROTEIN 1.
CC SITE 50 57 TONB_BOX.
```


FT SITE 895 912 TONE C-TERMINAL BOX
SQ SEQUENCE 912 AA; 103233 MW; 294C08991A652CC2 CRC64;
Query Match 3.9%; Score 158.5; DB 1; Length 912;
Best Local Similarity 18.9%; Pred. No. 0.0037;
Matches 194; Conservative 133; Mismatches 314; Indels 383; Gaps 51;
QY 5 TLKPI---VLSILLINTPLLAQAH---EPEQSVG-----LETVTVG-KSRP 44
DB 2 TKPYFRLSIISCLLSICVYKAEQTSIKDTKEAISSEVDQTOSTDESELETISVTAEKVRD 61
QY 45 RATSGLLHTSTADKII--SGDTRKAVNLGDALDGPVGHASQVGGGASAPV-IRGQT 101
DB 62 RKDNEV---TGLGKIKTSESISREQVNIIRDLYDFGIVSQVQGRGASSGYSTRGMD 117
QY 102 GRIKIVL-----NHHGETGMADFPDPAIMVDLTALSQQVEILRGP 142
DB 118 RNRVALLVDGLPQTQSVVQSPVARSYSGTGAINIEYENV-----KAVEISKGG 169
QY 143 VTLVSSGNVAGLV-----DVAQG-----KIPKMPENGVSSELGLRLSSGNLE 186
DB 170 SSSEYNGALAGSVTQSKSAADILEGDKSWGIGTQKNAYSXKNKGFTHSLAVAGKQGGFE 229
QY 187 KLTSGINIGLGNFVLHTEGLYKSGDYAVPYRNKLKLPDSPRFANGQHRAVLGWRK 246
DB 230 ---GLAIYTORNSI---ETQVHKDALKGVQSYNRLIAKFPENQSAFVWNEDECPKG--- 278
QY 247 RFRRYSRRDQYGLPAHSHVEDDC-----HADIWQKSLN-KRYL---OLYPHLLT 296
DB 279 ---YDECIPSAKPAILLSTKKTETVSVDYTGANRIKENPMK 316
QY 297 EEDVDYDNPGLSCGFHDDDDAHAAHNGKWDILRNKRYELRAEMKQPP------GF 348
DB 317 YESQSFWLRG---GYHFEQ---HYTGG---IFFTQKEDIR---DMTFPAYLSPTEKGD 365
QY 349 EALRVHLNRNDY---HHDEKAGDAVEN---FFNNOTQNAIRIELHOPIGRLKSGWQVYL 402
DB 366 LANRPYKQDYGAYQHIEDGRGVKYGASGLYFDEHHRKQV-----GLEVI 411
QY 403 GQKSSALSATSEAV---KQPMILDNKVQH-----YSFEGVE--- 435
DB 412 YENKKGAIIDKAVLSANQOQNIILDSYMRTHCSLYPNPSKNCRPTLDKPYYSRDRNV 471
QY 436 ---QANW-----DNFT----- 443
DB 472 YKEKHNMLQNLLEKKQQNWLTQIIVENLGFDFTSALQHKVLTFRVATADISDKTG 531
QY 444 ---LEGGVREKQKASIRYDKA---LIDRENY-----KQ 472
DB 532 KTRRNGRLREYPYLPKPKPYFAGEDHCNYQSSSNYRDCVKRLIKGKNYYFAARNNALG 591
QY 473 PLPDLG-----AHRQTA-----RSFALSGNWFYTPQHKLSIT--ASHQERLPST 514
DB 592 KYVDLGLGIRYVSRVKANESTLSVKFKNFWSWNTGLIVIKPTETWLDLSYRLSTGFRNPSP 651
QY 515 QELYAHGKHVATNTFFVGNKHLNKNERSNNIELALGYEGDRWQYNLALYRNFGNVI-YAQ 573
DB 652 AEMY--GWRYGGKNDVYVYKPKPEFSRQEFGLGALKGDFGNTIEISHFSNAYNLIAFAE 709
QY 574 TLN---DGRGPKSIDDESEKLVRYN-QSGADPYGAEIYF----- 611
DB 710 ELKNGTGRKNGYGHNAQNAKALGVNITAQLDPNGLWKLRIPIYGYWATFAYNRVKVKDQKI 769
QY 612 ---KPTPRYRIGVSGDYV-----RGLKNLPSLPGRDAYGN 645
DB 770 NAGLASVSSYLFDAIQFS-RYIIGLVYDHSNTWGNVATFTQSKAKSQNELLGKR-ALGN 827
QY 646 RPFIAQDQNAVPVPAARLGFHLKASLITRIDANLDYRVFAONKLARVETRPFPHMLN 705
DB 828 N---SRDVKSTRKLTRA---WHI-----LDVSGYWMANK-----NIMLR 860
QY 706 LGA-----NTR-----RNRTRYGEWNNYKADNLLNQSVIAHSSFLSDTFQMGSRFTGGV 754

Db 861 LGIYNLFNRYVTWVAVRQTAQAVNQH---QNVGNTRYAAS-----GRNYTLTL 908
QY 755 NVKF 758
Db 909 EMKF 912
RESULT 18
PFEA PSEAE
ID PFEA PSEAE STANDARD; PRT; 746 AA.
AC Q05038;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric enterobactin receptor precursor.
GN PFEA OR PA2688.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K407;
RX MEDLINE=93123148; PubMed=8419284;
RA Dean C.R., Poole K.;
RT "Cloning and characterization of the ferric enterobactin receptor
gene (pfeA) of Pseudomonas aeruginosa.";
RL J. Bacteriol. 175:317-324 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -I- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC
ENTEROBACTIN.
CC -I- SUBCELLULAR LOCATION: Outer membrane.
CC -I- INDUCTION: BY IRON AND ENTEROBACTIN.
CC -I- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; M38033; AAA25928.1; -;
CC EMBL; A5004697; AAG06076.1; -;
CC PIR; A40636; A40636.
CC HSSP; P05825; 1FEP.
CC InterPro; IPR000531; TonB boxC.
CC Pfam; PF00593; TonB dep Rec; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 25 POTENTIAL
FT CHAIN 26 746 FERRIC ENTEROBACTIN RECEPTOR.
FT SITE 39 44 TONB BOX.
FT SITE 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80967 MW; ADD4FB0CB23251 CRC64;
Query Match 3.9%; Score 155.5; DB 1; Length 746;
Best Local Similarity 19.3%; Pred. No. 0.0045;

QY 270 -----DDCHAD-----IIWQSLINKRYLOLYPHLITEED----- 299
 Db 364 LGKYSGNKAERLRFVQEGSTLOGIGYGTGVYDERHTKRYGYEYVYHNADKDTWADYA 423
 QY 300 -VDYDNPGLS-----CGFHDDDAHAHNGKPMIDLRNKR--YE-----LRAEWK 342
 Db 424 RLSVDRGIDLDNPLQTHCS-HGSDKNCRCPDGNKPYSFYKSDRMIEYBSRNLFOAVFK 482
 QY 343 QPPF-----GFALRVHLNRNDYH--HDEKAGDAV-----ENFFNNOTQN----- 380
 Db 483 KAFDTAKIRHNLINSINLGDFKSLQSHSDYVLQNAVOAYDLITPKPPFNGSKDNPRV 542
 QY 381 --ARIELRHOPIGRLKSGWGVYLGQKSSALSATSEAVKQPMLLDNKVQHYSPFGEVQAN 438
 Db 543 SIGKTTVNTSPICRFGNN-----TYDCTPRNITGNGGYAAV 579
 QY 439 WDNFTLEGGVRVERQKASIRYDKALIDRENYKQPLDGAHROTARSFALSGN----- 492
 Db 580 QDNVRLG---RWADVGAIRYDY----RSTHSEDKSVSTGTHRN-----LSWAGVVLK 626
 QY 493 ---WYFTPOHKLSTASHQERLPSTQELYAHGKHVAWTFEVGNKHLNKRNNIELALG 549
 Db 627 PFTWM-----DLTYRSTGTFRLPSFAEMYGWRAGESLKTLD-----LKPEKSFNREAGIV 676
 QY 550 YEGDRWQYNLALYNRFGNYI---YAQTLNDGR---GPKSIEDDSEWKLVRYNQSG-AD 601
 Db 677 FKDFGNLEASYFNNAIRDLIAFGYETRTQNGTSASGDPGYRNAQARIAGINILKID 736
 QY 602 FYGAEGLYFKPTPRYRIGVGDYVGRGLKNLPSLQREDAYGNRPPIAQ--DDQNAPRV 659
 Db 737 WHGVWGL---PDGLYS---TLAYNRKVKD-----ADIRADRTFTVSLFDVAVQPSR 783
 QY 660 PAARLGP-----HLKASLTDR-----IDANLDYRVVFAQNKLYARTYTRT 698
 Db 784 YVLGLGYDHPDGIMGINTMETYTSKAKSVDELGLSGALLNGANAKAASRRTRPWYVTDV 843
 QY 699 PG-----HMLNLGANYRNTYRGEN-NWYVKADNLNQ--SVYAHSSFLSDTPQWGRS 749
 Db 844 SGYINIKHLTRAGVYNLLNRYVTWENVRQTAGGAVNHQNVGVYNYAAP-----GRN 899
 QY 750 FTGGVNVKF 758
 Db 900 YTFSELMKF 908

RESULT 20

ID IUTA_ECOLI STANDARD; PRT; 732 AA.
 AC P14542;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Ferric aerobactin receptor precursor (Cloacin receptor).
 GN IUTA.
 OS Escherichia coli.
 OG Plasmid IncFI ColV3-K30.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=F344;
 RA Krone W.J.A., Stegehuus F., Koningsstein G., van Doorn C.,
 RA Roosendaal B., de Graaf F.K., Oudega B.;
 RT "Characterization of the pColV-K30 encoded cloacin DF13/aerobactin
 RT outer membrane receptor protein of Escherichia coli; isolation and
 RT purification of the protein and analysis of its nucleotide sequence
 RT and primary structure";
 RL FEMS Microbiol. Lett. 26:153-161 (1985).
 RN [2]
 RP REVISIONS.
 RA Oudega B.;
 RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.

CC --!- FUNCTION: RECEPTOR FOR CLOACIN DF13/AEROBACTIN.
 CC --!- SUBCELLULAR LOCATION: Outer membrane.
 CC --!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X05874; CAA29297.1; --
 DR EMBL; X05874; CAA29298.1; ALT INIT.
 DR InterPro; IPR00531; TonB boxC.
 DR Pfam; PF00593; TonB dep Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Plasmid;
 KW Receptor.
 FT SIGNAL 1 25
 FT CHAIN 26 732 FERRIC AEROBACTIN RECEPTOR.
 FT SITE 31 38 TONB BOX
 FT SITE 715 732 TONB C-TERMINAL BOX
 SQ SEQUENCE 732 AA; 81014 MW; 0C23879C0B27AE2B CRC64;
 Query Match 3.8%; Score 154; DB 1; Length 732;
 Best Local Similarity 19.9%; Pred. No. 0.0055;
 Matches 168; Conservative 109; Mismatches 301; Indels 266; Gaps 45;
 QY 6 LKPIVLSILLINPTLLAAAHETESQVLETVTVVVKSRPRA-----TSGLLHTSTASDKII 61
 Db 12 LNPILLTWM---APAVAQOTDDETFV---VSANRSNRTVAEMAQTWTWVIEABLEQIQ 64
 QY 62 SGTDLRQKAVNLGDLGVPCIHAS-----QYGGGA-SAPVIRQGTGRIRKVLNHHGTG 115
 Db 65 GGEKLDKALAQ-----IFGLVSSRSRTYGNVGRPLVLVDGVRLLNSTRSDTSRQL 118
 QY 116 DMAD-FSPDHAIMVDTALSOQVEILRGFVITLLYSSGNVAGLDVADGK-IPEKMPENGVS 173
 Db 119 DSIDPFNMWH-----IEVIFG-ATSLYGGSTGGLINIVTKGQPTWMEFEAG 166
 QY 174 GEIGLRLLSSGNLEKL-----TSGGINIGLGNFVLHTEG-----LYRKSQDYAVPRYRNKRL 226
 Db 167 TKSGFSSSKDHERIAGAVSGNE-----HISGRLSVAYQKFGW----- 206
 QY 227 PDGPRRFANGQRAVL-----GWRKFRYRTYSDRDQYGLPAHSHEYDDCHADIWOK 280
 Db 207 -----FDGNGDATLLDNTQTGLQ-----YSRDLIDMGT-----GTLNIDSRQL 245
 QY 281 SLINKRYLQLYPHLLTBEVDYDNPGLSCGFHDDDDAHAAHNGKPMID--LRNKRYELR 338
 Db 246 QLITQYY-----KSQGDYD---GLNLG---KGFSAIRGTSTPFVSNGLNSDR---- 287
 QY 339 AEWKQPPFPGEALRVHLNRND-----YHDE-----KAGDAVENFNN 376
 Db 288 -----IPGTDGHLISLQYSDSAFLGQELVGQVYRDESLRFYPPFTVANKQVTAFFSS 341
 QY 377 QTO---NARIELRHOPIGRLKSGWGVYLGQKSSA-----LSATSEAVKQPMLLD 423
 Db 342 QQTDQYGMKLTNLSKFPDQWQITWGLDAHERFTSNQMFDFDLAQASAGGLNNKNIYTT 401
 QY 424 NKVQHYVFFGVE---QANWD---NFTLEGGVRVEKQKASIRYDKALIDRENYKQPLD 477
 Db 402 GRYPYDITNLAAFLQSGYDINNLTFLNGGVRYQ-----YTNKIDDFIGYAAQ-RQI 453
 QY 478 GAHQTA-----RSFALSGN--WYFTPOHKLSTASHOERLPSTQELYAHGKHVA 525
 Db 454 GAGKATSADAFWRLSLRHLFNLNAGLMLHTEPCQAWLNFESQGLELPDPGKYVGRGIYGA 513
 QY 526 -----TNTFEVGNKHLNKRNNIELALGVEGDRWQYNLALYRNFRFGNYIAQ---T 574
 Db 514 AVNGHLP LTKSVNVVSDSKLEGVQVDSYELGWRFTGNLRTQIAAYYSISDKSVVANKDLT 573

Qy 575 LNDGRGPKSIEDSEMKLVRYNQSGADFYGAGEI-YFKPTPRYRIG-----VS 622
 Db 574 I-----SVDDKR-----IYGVGAVDYLIPDWDSTGVNENVLKTESKVN 615
 Qy 623 GDYVGRGLKNLPSLPGRDAY-----GNRPPIAQDDQ-----NA 656
 Db 616 GTWQKYDYKT--ASPSKATAYIGNAPDPWSLRVQSTTFDVSDAQGYKVDGYTTVDLLGS 673
 Qy 657 PRVPAARLGFHLKASLTDRIANLDYRVFAQNKLARLYETRPGHMLNL----- 706
 Db 674 YQLPVGTLFSFIE-NLFDK-----DYTWGQRAPLY--SPGPGASLDYKGRGRTF 724
 Qy 707 GANY 710
 Db 725 GLNY 728

RESULT 21
 FOXA_YEREN STANDARD; PRT; 710 AA.
 AC Q01674;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Ferrioxamine receptor precursor.
 GN FOXA.
 OS Yersinia enterocolitica.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=630;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=ATCC 51872 / WA-C / Serotype O:8;
 RX MEDLINE=92349959; PubMed=1640832;
 RA Baemler A.J., Hantke K.;
 RT "Ferrioxamine uptake in Yersinia enterocolitica: characterization of
 the receptor protein FoxA".
 RL Mol. Microbiol. 6:1309-1321(1992).
 RN [2]
 RP REVISIONS.
 RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
 RA Baemler A.J.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: FERRIOXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE
 TONB PROTEIN.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X60447; CAA42975.1; -.
 DR HSPF; P06971; IQQJ.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
 KW Transport; TonB box.
 FT SIGNAL 1 26
 FT CHAIN 27 710
 FT SITE 28 35
 FT SITE 693 710
 FT TRANSMEM 29 37
 FT TRANSMEM 65 73
 FT TRANSMEM 91 99
 FT TRANSMEM 106 114
 FT TRANSMEM 106 114

FT TRANSMEM 137 145
 FT TRANSMEM 152 160
 FT TRANSMEM 180 188
 FT TRANSMEM 194 202
 FT TRANSMEM 208 216
 FT TRANSMEM 259 267
 FT TRANSMEM 271 279
 FT TRANSMEM 293 301
 FT TRANSMEM 309 317
 FT TRANSMEM 353 361
 FT TRANSMEM 370 378
 FT TRANSMEM 427 435
 FT TRANSMEM 443 451
 FT TRANSMEM 476 484
 FT TRANSMEM 491 499
 FT TRANSMEM 517 525
 FT TRANSMEM 531 539
 FT TRANSMEM 555 563
 FT TRANSMEM 567 575
 FT TRANSMEM 579 587
 FT TRANSMEM 610 618
 FT TRANSMEM 624 632
 FT TRANSMEM 649 657
 FT TRANSMEM 671 679
 FT TRANSMEM 684 692
 FT TRANSMEM 702 710
 SQ SEQUENCE 710 AA; 78382 MW; 82EB6EC1546900B8 CRC64;

Query Match 3.8%; Score 151.5; DB 1; Length 710;
 Best Local Similarity 17.9%; Pred. No. 0.0079;
 Matches 144; Conservative 100; Mismatches 250; Indels 309; Gaps 36;

Qy 10 VLSILLINTPLLAQAHETEQSVGLETTVVVQKSRPRATSGLIHT----STASDK--IISG 63
 Db 12 ILCSLAMPFIPLASIAADDT-----IEVTAKAGHEADLPTSGYTATTTKGATKTDQPLILTA 66
 Qy 64 DTLR-----QKAVNLGALDGVPCGHASQYGGAS---APVIRGQGRRIKVLNHH 111
 Db 67 QSVSVVTRQQMDQNVATVQALNYPGVFTG-FSGGATRYDTVALRG-----FHG 116
 Qy 112 GETGD-----MADFPDHAIMVDTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKI 163
 Db 117 GDVNTFLDGLRLSDGGSYNVLQVDPWFLERIDVIRKGPSSALYQGSIPGGVV-----MM 171
 Qy 164 PERKPVNGVSGELGLRLSSGNLE-KLTSGGINIGLGNKFNVLHTEGLYRKSGDYAVPRYN 222
 Db 172 TSKRPQ--FTSEGHFLTAGNNNTQVAAPDVTDAISEHMAFLTGTITNSD----- 220
 Qy 223 LKRLPDSRPRPANGQHRVAVLQWRKRFYRTYSDDRDQGLPAHSHEYDDCHA--DIWQ 279
 Db 221 -----TMYDHOEREYAIAFSLLMQ 240
 Qy 280 ----KSLINKRYLQLYP----HLLTEEDVDYDNPGLSCGFHDDDDAHAAHNGKPWIDLR 331
 Db 241 PDENTSLLLRANLQKDPGGVHSAVPADGSIYGQKLSRGFFDGESNH----- 287
 Qy 332 NKRYELRAEWKQPPFGFEALRVHLNNDYHDEKAGDAVENFFNNQT---QVARIELRHQ 388
 Db 288 ----NVFKRWQIY-----SYEFSHKFPDD-VMSFRQNASYTHSNLTQLEQVYQ 329
 Qy 389 PIGRLKSGMGV-----QYLQCKSSALSA-----TSEAVKQPMILLDNKVQHSF 431
 Db 330 -----GGWNSDRITLMNRYISGEDSSLNFAFVNDQLEADLRTAAVKKHVL----- 374
 Qy 432 FGVEQANWDNFTLEGGVRVEKOKASIRYDKALIDRENYVKOPL-PDLGAH----- 480
 Db 375 -----GVDFOKFRNNLRSDSA-----YATPLNPTGVSGSGSTLYSYL 412
 Qy 481 -----ROTARFALSG-----NWY--FTPQHKLSLTASHOERLPSTOEL--- 517
 Db 413 LTTPTGINTSYLSRRYEQSGVYLQDEMTLDNHLNLSGRYDRMKTENINNTANSTDTDN 472
 Qy 518 YAHGKHVATNTFEVG-----NKHINKERSNNIELALGYE--GDRW 555

```

Db 473 HASGRASLLYSFDSGISPYVSQAITSLEPDAQOQLKLPWTSEQYEVGVIIYPPPGSTS 532
QY 556 QYNLALY---RNRFNGYIYAQTLDNGRGPKSIEDDSEMKLVRYNQSGADFYGABGEIYFK 612
Db 533 LYSAAVLDLTQNDVANRAVPATYVYPAG-----KVSQGL-----ELEAR 572
QY 613 PTPRYGVSGDYVGRGLKNLPSLPGREDAYGNRPFA----- 650
Db 573 SQISDRUSVIAGTYNVRVKFQDAIDGND---GNTFVLAPSNWASLWQAYEAGYGINVGAG 629
QY 651 -----QDDQNAAPRVPAAELG 665
Db 630 IRYIGKQWADDANTLRVPSYTLG 652

RESULT 22
TB11_NEIMO
ID TB11_NEIMO STANDARD; PRT; 915 AA.
AC Q01956;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TB11.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RX MEDLINE=92394880; PubMed=1325963;
RA Cornelissen C.N., Biswas G.D., Tsai J., Paruchuri D.K.,
RA Thompson S.A., Sparling P.F.;
RT "Gonococcal transferrin-binding protein 1 is required for transferrin
RT utilization and is homologous to TonB-dependent outer membrane
RT receptors.";
RL J. Bacteriol. 174:5788-5797 (1992).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M96731; AAA25503.1; -.
DR F1R; A43335; A43335.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 915 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB BOX.
FT SITE 898 915 TONB C-TERMINAL BOX.
SQ SEQUENCE 915 AA; 102213 MW; 697CF74B1010422F CRC64;

```

```

Query Match 3.7%; Score 150.5; DB 1; Length 915;
Best Local Similarity 18.4%; Pred. No. 0.013;
Matches 186; Conservative 132; Mismatches 344; Indels 349; Gaps 48;

QY 1 MAQTLKPIVLSILLITPLLAQHETE---QSVGLSTVTVGKS-----RPRATSGLLH 52
Db 1 MQQHLFRNLICLSLMTALPAYAENVQAOQAEKQLDITQVKAQKQKTRDRDNEVTGL-- 58

```

```

QY 53 TSTASDKII--SGDTL--ROKAVNLGDALDGVGPIHAGSYGGGASAPV--IRGQTGRRIKV-- 107
Db 59 -----GKLVKTADTLTSKEVLDIRDLTRYDPCIAVVEQGRASSGYSIRGDKNRVSLTV 113
QY 108 -----LNHHGETGDMADFPDPAHIMVDLTALSOQVEILRGPTVLLYSYG 150
Db 114 DGLAQIQSYTAQAALGTRTAGSSGAINEIEYENV-----KAVEISKGSNSVEQSG 165
QY 151 NVAGLV-----DVADGKIPEKMPENGSGELGRLRLSSGNLEKLTSGGINIGLKNEVL-- 203
Db 166 ALAGSAVAFQTKTADDVI-----GEGRWGIGQ-----SKTAYSGKRLTQSTALAGR 212
QY 204 --HTEGLYRKSGDYA-----VPRVNLKRLPDSR-----RPAQCQ--- 238
Db 213 IGGAEALLIIRTAGHAGEIRAHEAAGRGVQSFNRLAPVDDGSKYAYFIVEECKNGHEKC 272
QY 239 -----RAVLG-----W-----RKEFY-----R 250
Db 273 KANPKQVVGEDKQTVSTRDYTGPNRFLADPLSYESRSWLFPRGFRFENKRHYIGGILE 332
QY 251 RTYS--DRRDQ-----YGLPAHSHEYD-----D 271
Db 333 RTQQTFTDTRDMTVPAFLTKAVEDANQKQAGSLRGNGKYAGNHKGGLFTSGENNAVPAE 392
QY 272 CHADIIWQKSLINKRYLQLYPHLLTEED-----VDYDNPLGSCGFH-----DDDD 316
Db 393 YGTGVFYDETHTKRGYLEVYVYTNADKDTWADYARLSYDRQIGLDNHFQOHTHCADGSD 452
QY 317 AHAHAHNGKPMIDLRNKRY-----ELRASWKQFPF-----GPEALRVHLNR 357
Db 453 KYCPASADKPFYSYKSRVYVYGESHKLLQAAFKKSDTAKIRHNLVNLGDRFQSGNLRH 512
QY 358 NDYHHDEKAGDAVENFFNNQTONARIELRHQPIGRLKGSWGVQVYLGQKSSALSASEAVK 417
Db 513 QDYVY-----QSANRAYSLKTPQNNNGKTSPPNGREKNPYV-----SIGRGNVYTR 559
QY 418 QPMLLDN-----KVQHYSPFGVEQANWDNFTLEGGRVEKQKASIRVDKALIDREN 468
Db 560 QICLFGNNTYDCTPRSRNGKSYAAVR---DNVRLG---RWADVGAGLRVD----- 605
QY 469 YYKQPLDPLGA--HQRTARSPALSGNWTFTPOHKLSLT--ASHQBELPSTQELYAHGKHA 525
Db 606 -YRTHSDDGVSVTGTHRTLSWNAGIVLKPADWLDLTYRTSTGFRLPSPFAEMYG----- 658
QY 526 TMTFVGNK-----HLNKRSSNNIELALGYEGDRMQYNLALYRNRFNGYI---YAQTLDNG 578
Db 659 ---WRSGDKIKAVKIDPEKSPENKAGIVFKGDFGNLRASFNNAYRDLIVRGYEAQIKDG 715
QY 579 R-----GPKSTIEDDSEM-----KL-----VRYNOSGADF 602
Db 716 KEQVKGMPAYLNAQSARITGINILGKIDWNGWDKLPFGWYSTTFAYNRVVRVDIKKRADR 775
QY 603 YGARGEIY--FKPT-----PRYRIGVSGDYVVRGRLLKNLPSLPQREDAYGNRPFA 650
Db 776 TDIQSHLFDALQPSRYVVGSGYDQPEGKVGWNGMLTYSKAKEITEL-----LGSRLIN 829
QY 651 QDDQNAAPRVPAAELGPHLKLASLTDRIANDLDYRVFPAQNKRLARYETRTPGHHMLLGANY 710
Db 830 GNSRNTKATARRTRPWI-----VDVS--GYITV-----KKHFTLRAGVYN 868
QY 711 RENTRYGEW--NYYKADNLNQ--SVVAHSFSLDTPQMGSRSTGCVNVKPF 758
Db 869 LLNHRVYTWENVROTAAAGAVNQHKNGVYNYRAAP-----GRNYTFSLEMKF 915

RESULT 23
TB11_NEIMB
ID TB11_NEIMB STANDARD; PRT; 911 AA.
AC Q09056;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.

```

GN TBPI.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=CCUG 37608 / M982 / Serogroup B / Serotype 9;
 RX MEDLINE=93345825; PubMed=8344530;
 RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
 RA Jacobs E., Schryvers A.B.;
 RA "Cloning and characterization of Neisseria meningitidis genes
 RT encoding the transferrin-binding proteins Tbp1 and Tbp2.";
 RL Gene 130:73-80(1993).
 CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
 CC TRANSFERRIN UTILIZATION.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- INDUCTION: By iron starvation.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z15130; CAAY78833.1; -;
 DR PIR; JN0821; JN0821.
 DR InterPro; IPR00531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Receptor; Signal; TonB box.
 FT SIGNAL 1 24
 FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.
 FT SITE 38 45 TONB BOX.
 FT SITE 894 911 TONB C-TERMINAL BOX.
 FT SITE 911 911 MW; 99283ABAE0B773B6 CRC64;
 SQ SEQUENCE 911 AA; 101631 MW; 99283ABAE0B773B6 CRC64;

 Query Match 3.6%; Score 144.5; DB 1; Length 911;
 Best Local Similarity 19.0%; Pred. No. 0.035;
 Matches 191; Conservative 135; Mismatches 337; Indels 343; Gaps 52;

 QY 1 MAQTLLKPIVLISILLINTPLLAQHETE-----QSVGLETVTVVVGKS-----RPRATSGLLH 52
 Db 1 MQQOHLFRNLCLSLMTALPAVAENVQAOAQEKQDITQVAKKQKTRDNEVTGL-- 58
 QY 53 TSTASDKII-SGDTL-RQKAVNLGDALDGVPGTHASQYGGASAPV-IRGQTRRIKV-- 107
 Db 59 -----GKLVTADTLSEKQVLDRLDTRYPDGIADVVEQGRGASSYIRGMKNRVS LTV 113
 QY 108 -----LNHHGETGDMADFPDPAHIMVDTALSSQVFEILRGVPVTLVYSSG 150
 Db 114 DGLAQISYTAQAALGTRTAGSSGAINELIENV-----XAVEISKGSNVEQSG 165
 QY 151 NVAGLV-----DVADGKIPEK-----MPENGVSSELGLRLSSGNLEKLTSGNIGLGNKF 201
 Db 166 ALAGSVAFOTKTADDVIGEGROWGIGQSKTAYSGK-----NRGLTQSIALAG-RIGGAEL 219
 QY 202 VLHTEGLYKRSQDY-----AVPRYNLKRLPDS----- 229
 Db 220 LIHTG--RRAGEIRAHEDAGRGVQSFNRLVPVEDSEVAYFIVECEGKNVETCKSKP 276
 QY 230 -----PRRF-----ANGQRAVL-----GMR-----KRFY-----RR 251
 Db 277 KKDVGKDERQTVSTRDYTGPNFLADPLSYESRSLFRPGFFENKRRHYIGILEHTQQ 336
 QY 252 TVSDRRD-----QVG-IP-----AHSHEYDDCHAD----- 275
 Db 337 TF-DTRDMTVPAFLTKAVFDANSKQAGSLPGNGKYAGNHKYGGLFTNGENGALVGAEGYT 395

QY 276 -IIWQSLINKRYLQLYPHLLTEED-----VDYDNPGLSCGFH-----DDDDAHA 319
 Db 396 GVFDYDETHKSRVLEVVYTNADKOTWADYARLSYDRQIGLNDHFQOQTHCSADGSKYC 455
 QY 320 HAHNGKPDIDLNKRY-----ELRAEWKQPPP-----GFEALRVHLNRNDY 360
 Db 456 RPSADKPFYSYKSDRVYIGESHRLLOAFAFKGFDATAKIRHNLVNLGFORFDSNLRHODY 515
 QY 361 HDEKAGDAVENFFNNQTONARIELRHQIPG-RLKGSW-----GVQYLQ-----KSSAL 409
 Db 516 YY-QHANRAYSSKTPPKTAN-----PNGDKSKPYWVSIGGNVVTGQICLFGNNTYT 566
 QY 410 SATSEAVKQPMLLDNKVQHYSFEGVEQANWDNFTLEGVVRVEKQKASIRYDKALIDRENY 469
 Db 567 DCTPSI-----NGKSYIA-----AVRDNVRVG--RWADVAGLRYD----- 601
 QY 470 YKQPLDPLCA-HROTARSPALSNGWYFTPOHKLST--ASHQERLPSTOELVYANGKHVAT 526
 Db 602 YRSTHSDGGSVSTGTHRTLSNAGIVLKPADMLDLYTSTGFRLPFAEYMGWRSGVQS 661
 QY 527 NTFEVGNKHLNERSNIELALGYEDGRWQYNLALYRNRFGNYIY---AQTNL---DGR 579
 Db 662 KAVKI-----DPEKSNKEAGIVFKGDFGNLEASWNNAYRDLIVRGYEAQIKNGKEAK 716
 QY 580 GPKSIDDDSEMKLVRYNQSG-----ADFYGAEG 607
 Db 717 GDPAYLNAQASARITGINILGKIDWNGVDKPEGWYSTFAYNRVHVDRDIKRAADTDIQS 776
 QY 608 ELYFKPTP-RYRIGVSGDYVRG-----RLKNLPSLPQREDAYGNRPPIAODQN 655
 Db 777 HLFDAIQPSRYVYVGLGYDQPEGKWGVNGMLTYSKAKEITEL-----LGSALLNGSRN 830
 QY 656 APRVPAARLGFHLKASLTDRIDANLDYRVVFAQNKCLARYETRTPGHMLNLGANYRNRTR 715
 Db 831 TKATARRTPWYI-----VDVS-CYITI-----KKFTLRAGYVNLNLYR 869
 QY 716 YGEW-NWYVKAADNLLNQ--SVYAHSSFLSDTPQMGSRFTGGVNVKF 758
 Db 870 YVTWENVROTAGAVNQHNKGVYNYRAAP-----GRNYTFSLEMKF 911

 RESULT 24
 FHUE_ECOLI
 IN FHUE_ECOLI STANDARD; PRT; 729 AA.
 AC P16869; P77292;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE FhuE receptor precursor (Outer-membrane receptor for Fe(III)-coproten,
 DE Fe(III)-ferrioxamine B and Fe(III)-rhodotruilic acid).
 GN FHUE OR B1102.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90286919; PubMed=2162465;
 RA Sauer U., Hantke K., Braun V.;
 RA "Sequence of the fhuE outer-membrane receptor gene of Escherichia
 RT coli K12 and properties of mutants.";
 RL Mol. Microbiol. 4:427-437(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).
 RL

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X69214; CAA49148.1; --
 DR EMBL; AE002504; AAF41895.1; --
 DR PIR; G81070; G81070.
 DR TIGR; NMB1540; --
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_dep_Reg; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
 KW Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 943
 FT SITE 826 943
 FT LACTOFERRIN BINDING PROTEIN A.
 FT TONB C-TERMINAL BOX.
 FT P -> Q (IN REF. 1).
 FT IAT -> VAA (IN REF. 1).
 FT A -> S (IN REF. 1).
 FT QAGGAT -> NPETA (IN REF. 1).
 FT I -> V (IN REF. 1).
 FT V -> A (IN REF. 1).
 FT R -> H (IN REF. 1).
 FT E -> A (IN REF. 1).
 FT D -> N (IN REF. 1).
 FT DIKRTREPPFSV -> GIKKPSGGVEFLA (IN REF. 1).
 FT RES -> SEL (IN REF. 1).
 FT L -> V (IN REF. 1).
 FT YGK -> NGN (IN REF. 1).
 FT Q -> M (IN REF. 1).
 FT E -> K (IN REF. 1).
 FT KNLV -> OKLI (IN REF. 1).
 FT A -> K (IN REF. 1).
 FT K -> N (IN REF. 1).
 FT STGFENNQ -> YSDYTDKG (IN REF. 1).
 FT L -> V (IN REF. 1).
 FT V -> L (IN REF. 1).
 FT SEQUENCE 943 AA; 105680 MW; BD569ECACFC01A84 CRC64;

Query Match 3.4%; Score 138.5; DB 1; Length 943;
 Best Local Similarity 19.4%; Pred. No. 0.096;
 Matches 200; Conservative 126; Mismatches 328; Indels 379; Gaps 54;

QY 8 PIVLSILLINTPLAAQA-----HETEQSVGLETVTV---VCKSRPRATSGLLHTSTA 56
 DB 8 PLTLTALAIAATAPAYAAQAGGATPDAATQSLKEITVRAAKVGR-RSEKATGL----- 60
 QY 57 SDKII-SGDTL-RQKAVNLGDALGVFGIHAQYGGGASAPV-IRGQTGRRIKV----- 107
 DB 61 -GKIVKTSETLNKEQVLGIRDLTRYDGVAVVEQNGASGGYSIRGVDKRVAIVSDGVA 119
 QY 108 -----LNHHGETGDMADFPDRAIMVDTLSQVEILRGVILLVSSGNVAGLVD 157
 DB 120 QIOAFTVQSGSLGYGGGGGAINIEYENIST-----VEIDKAGSDHSGSALGGAVA 174
 QY 158 -----VADGK---IPEKMPENGVSGE-----LGLRLSSGNLEK----- 188
 DB 175 FRTKEADLISDGKSWGIGQAKTAVGSKNRQPMKSLGAGFKDGEGLLITEROGRETRP 234
 QY 189 ---TSGGINTGLGK-NFVLHTEGLYRKS-----GDY----- 215
 DB 235 HGDIAADVGYGIDRLDAFQTYDIKTRTREPFFSVGEERESKPVAKLAGYKYLNNQLNR 294
 QY 216 -----AVPRYRL-----KRLDPSRPRFANGQRAVLGV 244
 DB 295 WVKERIQNPLSABEEAQVREQAQRHENLSAQAYTGGRILPD-PMDYRSGSWLAKIGY 353

QY 245 R---KRFVRRYTSRRDOYGLPAHSHEYDDCHADIIWOKSLINKKYLQLYPHLLTEEDVD 301
 DB 354 RFGGRHYVGGVFEDTKORYDI-----RDMTEKQYVG-----TDEAEK 390
 QY 302 YDNFGLSCGFHDDDDAHAAH---NGKPW-----IDLNRKEVELRA 339
 DB 391 FRD---KSGVYDGDGDFRDGLYFVNIEEWKGDKNLVRIGLKYKSYTKFIDEHRRRRMGL 447
 QY 340 EWQFFQFEALRVLHRLNDYHDEKAGDAVENFNN---OTQNARIEL----- 385
 DB 448 LYRY-----ENEAYSNDWADKAVLSPDKQGVATDNTLKNCAVYPADVXSC 494
 QY 386 -----RHQPIG-----RLKSGWGVQY---LQKSSALSATSEAVKQ 418
 DB 495 RASADKPYSDSDRPHYREQHNVLNASFESLKNKNTKHLTLGLFGYDA---SKAISR 550
 QY 419 PMLLDNKKVQHYV-FFGVQOANDNFTI---EGGV-----RVEKQKASIR----- 458
 DB 551 PEQLSHNAARISESTGFDENNQDKYLLGKPEVSGVGYIETLRSKCVPRKINGSNIH 610
 QY 459 -----YDKAL---IDRENYKQPLDLAGAHRQTARSFALSG-----NW---Y 494
 DB 611 ISLNDREFSIGKYFDFSLGGRYDRNF-----TTSEELVRSRGYVDRSWNSGIL 658
 QY 495 FTPOHKLST---ASHQERLPSTQELYAHGKIVATNTEVFG---NKLHKNERNNIELALGY 550
 DB 659 FKPNRHFSVSYRASSGFRTEFSQELFGIDY---HDYKPGWQRPALKSEKAAANREIGLQW 715
 QY 551 EGDREWYNLALYRNFRGNYIVAQTLNDGRGPKSIEDOSEMKLVRYNQSG-----ADFY 603
 DB 716 KGDFGLFLEISFRNRYTDMIAVA-----DHKTCLP---NQAGQLTEIDIRDY 760
 QY 604 GAE-----GEIYFKPTPRYRIGVSGDYVRGLKNLPSLPQREDAYGNRPFAQDD 653
 DB 761 NAQWMSLQGVNLIKIDWN-----GVYKLPGLYTLTAYNRKPKSVNRPGLSLRS 813
 QY 654 QNAPRVPAAR---LGF-----HLKASLTDRIANDLYYRVAQNKLARYETRTPGHH 702
 DB 814 YALDAVQPSRYVLGFGYQDPGKNGANIMLTYSKGNPDELAAYLAGDQ-KRYSTKRASS 872
 QY 703 -----MLNL-----GANRY-RNTRYGEW-NWYVKADNLLNQSVYAHSSFLSDTPQ 745
 DB 873 WSTADVSAYLNLKRLTLRAAIYGNVRYVTWESLQTAESTANR---HGGDSNRYGYAA 930
 QY 746 MGRSFTGGVNVKF 758
 DB 931 PGRNFSLALEMKF 943
 RESULT 27
 ID FFTA_PSEAE STANDARD; PRT; 720 AA.
 AC P42512;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Fe(III)-pyochelin receptor precursor.
 GN FFTA OR P44221.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OC NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-51.
 RC STRAIN=PAO / IA602;
 RX MEDLINE=94117363; PubMed=8288523;
 RA Akenbauer R.G., Quan H.N.;
 RT "Fpta, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a
 RT phenolate siderophore receptor homologous to hydroxamate siderophore
 RT receptors." J. Bacteriol. 176:307-319 (1994).
 RL [2]
 RP SEQUENCE FROM N.A.

DR PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE_NEG.
 DR PROSITE; PS01156; TONB DEPENDENT REC 2; 1.
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor.
 FT SIGNAL 1 45 POTENTIAL.
 FT CHAIN 46 809 FERRIC-PSEUDOBACTIN BN7/BN8 RECEPTOR.
 FT DOMAIN 156 160 POLY-SER.
 FT SITE 732 809 TONB C-TERMINAL BOX.
 SQ SEQUENCE 809 AA; 88389 MW; 0B339F6E788A8C0D CRC64;
 Query Match 3.3%; Score 132; DB 1; Length 809;
 Best Local Similarity 18.9%; Pred. No. 0.22;
 Matches 133; Conservative 98; Mismatches 247; Indels 224; Gaps 34;
 QY 11 LSLILINTPLLA-----QAHETEQSGLVETVTVVVKSRPRATSG-LHTSTASDK 59
 DB 100 LAILLAGTGLASRGANASYSLOASASTGALSASVSGKAPGSTTGTGLITYTSSS 159
 QY 60 -----IISGDTL-RQKAVNLGDAIDGVPGIHASQYGGGASAPVIRGQT 102
 DB 160 STRLNLTPRETQSLTWTRQLDDQLRLNTDALEATPGITVVRDGLGSES- ---SYMS 216
 QY 103 RRIKVLNHGE---TGDMAFPDHAIMVDTALSOQVELIRGPFVTLVSSGNVAGLVDA 159
 DB 217 RFAIQNYEVDGVPTSTRLDNYQSWMAFD-----RVELVRGATGLISGMGNPSATINLI 271
 QY 160 DGKIPKMPENGVSGLGRLSSGNLEKLTSGGINIGLKNFVLHTEGLYRKS- --- 213
 DB 272 -RKRPATAQAQITGE-----AGNWDRYGTG-----FDVSGPLTETGTNIRGRF 313
 QY 214 --DYA-----VPRYNKLKLP-----DS-----PRRAN 235
 DB 314 VADYKTEKAWIDRYNQOSQMTGITEFDLSEDTLTVGFSYLRSDIDSPLSGLTRFST 373
 QY 236 GQH-----RAVLGKRKFFYRT--YSRRDQYGLPAHSHEYDDCHADIIWQKSLINKR 286
 DB 374 GERTNLKRSINAAPDWSYNDHEQTSFPTSIEQLG-NWMSGKIELTHAE-----NK 423
 QY 287 YIQLYPHLITEEDVDYDNGLS-----CGFHDDDAHAHAHNGKFWIDLNRKRYEL--- 337
 DB 424 FDELNFAMGE--LNPDPGSLGSLQPLVRFSGTFRQNDLIDYATGP--FSLGFRHELITG 478
 QY 338 --RAEWKQPPGPEALRVHLNRDHYHDEKAGDAVENFNQNTQNAIRLHQPIRLGLK 395
 DB 479 MTLQSVRENTPSWGNR-----YDYAGSPAGIDNLFNWDGSAK-----PAFVBSG 525
 QY 396 SWGV---QYLGQKSSALSATSEAVKQPMLLDNKVQHSYFPGVEQAWNDFTLE---GGVR 449
 DB 526 KSSIDEDQYAAYLTSRFSVTDDL---SLILGSL-----INWKDTSDRPYGSEE 572
 QY 450 VEKQKASIRYDKALIDRENYKQPLDGAHRQTARSFALSGNWYFTPO----- 498
 DB 573 TEVNR-----EENGVFIPYAGVGYDLDDTWLSLYASYTKIPNPOGAWVTDESINKP 621
 QY 499 -----HKLSLTASH-----QERLPSTQ--ELYAHGKHVAATNTEFVGN 533
 DB 622 LDPMEGVGYELGKTHLGNKLNSSLAVFKLEQDNLAIWQHDNVYSAEQDTT-SKGIEL-- 679
 QY 534 KHLNERSNNIELAGY-----EGDRWQYNLALYENRGNY 569
 DB 680 -ELNGELAEQWASAGYSVVTDDADDQRIINTLNLRNSFKTF 720
 RESULT 29
 HTPG VIBCH
 ID HTPG VIBCH STANDARD; PRT; 635 AA.
 AC P23259; Q9KTB8;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chapterone protein htpg (Heat shock protein htpg) (High temperature protein G).
 GN HTPG OR VC0985.
 OS Vibrio cholerae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN 111
 RC SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 Ermolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi I., Sellers P.,
 McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Wenter J.C.,
 Fraser C.M.;
 RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae.";
 RT Nature 406:477-483 (2000).
 RL [2]
 RN SEQUENCE OF 1-110 FROM N.A.
 RP STRAIN=El Tor E7946;
 RC MEDLINE=91088618; PubMed=2124707;
 RX Parsot C., Mekalanos J.J.;
 RA "Expression of ToxR, the transcriptional activator of the virulence
 factors in Vibrio cholerae, is modulated by the heat shock
 response.";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:9898-9902 (1990).
 CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
 similarity).
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- INDUCTION: BY INCREASE IN TEMPERATURE
 CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 DR EMBL; AE004180; AAF94146.1; -.
 DR EMBL; M58033; AAA27574.1; ALT_INIT.
 DP PIR; B82255; B82255.
 DA HSSP; P07900; 1YER.
 DR TIGR; VC0985; -.
 DR HAMAP; MF_00505; -; 1.
 DR InterPro; IPR003594; AtPbind_ATPase.
 DR InterPro; IPR001404; Hsp90.
 DR Pfam; PF02518; HATPase_C; 1.
 DR Pfam; PF00183; HSP90; 2.
 DR PRINTS; PR00775; HEATSHOCK90.
 DR SMART; SM00387; HATPase_c; 1.
 DR PROSITE; PS00298; HSP90; 1.
 KW Chapterone; ATP-binding; Heat shock; Complete proteome.
 FT DOMAIN 1 344 A; SUBSTRATE-BINDING (BY SIMILARITY).
 FT DOMAIN 345 561 B (BY SIMILARITY).
 FT DOMAIN 562 635 C.
 SQ SEQUENCE 635 AA; 72197 MW; A635C6A38C151AAC CRC64;
 Query Match 3.2%; Score 131; DB 1; Length 635;
 Best Local Similarity 18.8%; Pred. No. 0.18;
 Matches 154; Conservative 95; Mismatches 282; Indels 286; Gaps 40;
 QY 34 ETVTVVKSR--PRATSGLLHT---STASDKIISGDTLRQKAVNLGDAIDGVPGIHASQY 88
 DB 3 ETATNTKRTGRFQSEVKQLLHMTHTSLYSNKEI---FLRELISNASDAVDKL----- 51
 QY 89 GGGASAPVIRGQTGRRIKVLNHGEGTGMADFSPDHAIMVDTALSOQVELIRGPFVTLVLS 148
 DB 52 -----RFOALSH-----PD-LYQGDALGVKLSFDKQNTLTIS 84
 QY 149 SGNVAGLVDAVDGKIPEKMPENGVSGLGRLSSGNLE---KLTS-----GGINIG 196

Db 668 LIASRAVDAETKGVDELSEVLPGWNVFTGYSHTR-----TEDADGHR----- 712
Qy 651 QDDQNAPRVPARLGFHLKASLTDRIDANLDYRVFAQNKLYARXTFTFGH-HMLNLGAN 709
Db 713 ----LTPQLP-----MDTFRFWNTYRL-----PGWEKLTLCGG 742
Qy 710 YRRNTR-----YGEWNYVYKADNLLNQSYVA 735
Db 743 VNWSKSTLNFARNYNSHVTQDDYFVTSLMA 772

RESULT 31
ID GTFB_STRMU
AC P08987; 069381; 069384; 069387; 069390; 069396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN GTFB OR SMU.1004.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacilliales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Veda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype E, MT4251 / Serotype F,
RC MT4467 / Serotype E, and MT8148 / Serotype C;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: Sucrose + {(1,6)-alpha-D-glucosyl}(N) = D-
CC fructose + {(1,6)-alpha-D-glucosyl}(N+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DISEASE: DENTAL CARIES.
CC -!- MISCELLANEOUS: GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
CC 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
CC WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH
CC FORMS OF GLUCANS.
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S.MUTANS.
CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17361; AAA88588.1; -
DR EMBL; D88651; BAA26101.1; -
DR EMBL; D88654; BAA26105.1; -
DR EMBL; D88657; BAA26109.1; -
DR EMBL; D88660; BAA26113.1; -
DR EMBL; D89977; BAA26119.1; -
DR EMBL; AE014940; AAN58705.1; -
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries;
KW Complete proteome.
FT SIGNAL 1 34
FT CHAIN 35 1476
FT DOMAIN 35 1051
FT DOMAIN 1097 1476
FT REPEAT 1097 1130
FT REPEAT 1161 1470
FT REPEAT 1161 1210
FT REPEAT 1225 1275
FT REPEAT 1290 1340
FT REPEAT 1355 1405
FT REPEAT 1420 1470
FT VARIANT 62 62
FT VARIANT 65 65
FT VARIANT 68 68
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 957
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1264 1264
FT VARIANT 1272 1272
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT CONFLICT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310

POTENTIAL.
GLUCOSYLTRANSFERASE-I.
CATALYTIC (APPROXIMATE).
GLUCAN-BINDING (APPROXIMATE).
A REPEAT.
5 X TANDEM REPEATS.
1.
2.
3.
4.
5.
S -> T (IN STRAIN MT4239).
T -> I (IN STRAIN GS-5).
V -> A (IN STRAINS GS-5, MT4245, MT4251,
MT4467 AND MT8148).
Q -> P (IN STRAIN MT4251).
I -> S (IN STRAINS GS-5, MT4245, MT4251,
MT4467 AND MT8148).
S -> F (IN STRAIN MT4251).
K -> N (IN STRAIN MT4251).
S -> D (IN STRAINS GS-5, MT4467 AND
MT8148).
N -> R (IN STRAIN MT4239).
I -> T (IN STRAIN MT4239).
K -> R (IN STRAIN MT8148).
F -> Y (IN STRAIN MT8148).
T -> I (IN STRAIN MT8148).
A -> V (IN STRAIN MT8148).
F -> L (IN STRAIN MT8148).
FGKPEV -> YGTPVA (IN STRAINS GS-5, MT4239
AND MT4467).
SV -> NT (IN STRAINS GS-5, MT4239 AND
MT4467).
ADS -> VDG (IN STRAINS GS-5, MT4239 AND
MT4467).
A -> T (IN STRAIN MT4239).
S -> N (IN STRAIN MT4239).
H -> Y (IN STRAIN MT4251).
E -> K (IN STRAIN MT8148).
Y -> C (IN STRAIN MT8148).
A -> P (IN STRAIN MT4239).
R -> P (IN STRAIN MT8148).
Y -> H (IN STRAINS GS-5 AND MT4467).
AND MT8148).
S -> G (IN STRAINS GS-5, MT4239, MT4467
AND MT8148).
H -> Y (IN STRAINS GS-5 AND MT4467).
Y -> H (IN STRAINS GS-5, MT4239, MT4467
AND MT8148).
S -> G (IN STRAINS GS-5, MT4239, MT4467
AND MT8148).
Y -> H (IN STRAIN MT4467).
R -> A (IN REF. 1).
ADQVRAASTAPSTDGK -> LKMFALRLARPHQMA
(IN REF. 1).
H -> L (IN REF. 1).

RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of *Neisseria*
RT meningitidis 22491.";
RL Nature 404:502-506(2000).
CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL162757; CAB84967.1; --
DR PIR: C81798; C81798.
DR InterPro: IPR000531; TonB boxC.
DR Pfam: PF00593; TonB dep Rec; 1.
DR PROSITE: PS00430; TONB DEPENDENT REC 1; FALSE_NEG.
DR PROSITE: PS01156; TONB DEPENDENT REC 2; 1.
DR Outer membrane; Iron transport; Transceptor; TonB box; Signal; Receptor;
KW Complete proteome. 1 27
FT SIGNAL
FT CHAIN 28 944 LACTOFERRIN BINDING PROTEIN A.
FT SITE 827 944 TONB C-TERMINAL BOX.
FT SSEQ 944 AA; 105732 MW; CE06B6192E74AE3E CRC64;
SQ
Query Match 3.2%; Score 127.5; DB 1; Length 944;
Best Local Similarity 19.7%; Pred. No. 0.56;
Matches 197; Conservative 117; Mismatches 335; Indels 351; Gaps 54;
QY 24 AHETQSQVLETVTV---VGKSRPRATSGLLHTSTASDKII--SGDTL--RQKAVNLGDAL 77
DB 31 ASDAASQSLKEITVRAAKVGR-RSKATGL-----GKIVKTSETLNKEQVLGIRDLT 82
QY 78 DGVPGIHASQVGGGASAPV-IRGQTGRRIKV-----LNHHGETGDMADF 120
DB 83 RYDPGVAVVEQNGASGGSYIRGVDKNRVAVSDVGAQIQAFVQGSLSGVGGGSGAI 142
QY 121 SPDHAIWDTALSOQVEITLPGVTLVLYSSGNVAGLVD-----VADGK---IPKMP 168
DB 143 NEIEYENIST-----VEIDKAGSDHGSGLGALGAVAFRTKBAADLSDGKSMGIQAKTA 197
QY 169 ENGVSGE---LGLRLSSGNLEKL-----TSGGINTGL-----197
DB 198 YGSKNRQFMKSLGAGFSKDGWEGLLIRTERQGRETRPHGDIADGVEYIDRLDAFRQTYD 257
QY 198 -----GKNFVLHTE-----GLY-----RKSQDY-----215
DB 258 IOKQNKAEYFLAEGEREKPKVAKLAGNLYLKNQLNRWVEERKKNQPLNAEEAAMVRE 317
QY 216 AVPRYRNL-----KRLPDSRRFRANGOHRAVLGWR---KRFYRTTSDRRDQGLP 263
DB 318 AQARHENLSAQSYTGGRILPD-PMDYRSGSLAKLGLYRGGRHYGVGVFEDTKQRYDI- 375
QY 264 AHSHEYDDCHADIIWQKSLINKRYLQYLPHLLTEEDVDYDNPGLSCGFHDDDAHAHAHN 323
DB 376 -----RDMTEKQYIG-----TDEAKKFSN-----KSGVYDGN-----402
QY 324 GKPWIDLRNKRYELR--AEWKQPPFGFEALRVHLNRN---DYHH-----362
DB 403 -----DFRDGLYFVFNIEEWKGDYTNLVKGIKLYSRTKFIDEHRRRRMGLLYRENEAY 457
QY 363 -DEKAGDAVENFFNN--QTONARIEL-----RHS 387
DB 458 SDNWADKAVLSFDKQGVATDNNTLKLNCVAVYPAVDKSCASADKPKFSYDSSDRFVREQH 517
RA
SQ SEQUENCE 1476 AA; 165846 MW; 9C6E09F731B4CBFC CRC64;
Query Match 3.2%; Score 129; DB 1; Length 1476;
Best Local Similarity 20.6%; Pred. No. 0.84;
Matches 144; Conservative 89; Mismatches 236; Indels 230; Gaps 44;
QY 87 QYGGASAPVIRGQTRRTKVLN--HHGETGMDAFSDHAIWV---DTALSOQVEILRG 141
DB 895 RYDLGISKENKYGTDADLKAIKALHSGIKVMADWDFDQYAFPEKEVVTATVDRFGK 954
QY 142 PV-----TLLY-----SSGN-----VAGLVDDVADGKIPE-----KMPENGVSSELGLRL 180
DB 955 PVEGSIQKSVLYVADSSKSGDQAKYGGAFLEELQAKYPELFARKQISTGVPMDSVKI 1014
QY 181 SSGNLEKLTSGGINIGLKNFVLHTEGLYKSGDYVAPRYNL---KRLPDSRRFRANGQ 237
DB 1015 KOWS-AKYFNGTNILGRGAGYVLK-----DOATNTYFNISDNKEINFLPTLLNQD 1064
QY 238 HRAVLGWRKRFY-----RRTYSRRDQYGLPAHSHEVDCHADIIWQKSLINKRY 287
DB 1065 SQVGSYDYGKGVYVYSTGYQAKNTFISBGDKW-----YVFDNNGYVMTGAQINGVNY 1118
QY 288 -----LQYLPHLLTEEDVDYDNPGLSCGFHDDDDAHAAHNGKFWIDLRNKRYE-----336
DB 1119 YFLNSGLQLRDAILKNEDGY-----AYYGN-----GRRYENGYYQ 1155
QY 337 -LPAEWKQPPGFEALRVHLNRD-----YHHDE-----KAGDAVENF 373
DB 1156 FMSGVWR-----HFNNGEVSGLTVIDGQVQFDEMGYQAKGKFTTADGKIRY 1204
QY 374 FNNQTONARIELRHOPIGRLKSGWQVYLQKSSALSATSAAVKPMLDNKVOHYFFG 433
DB 1205 FDKQSGNM---YRNRFFIENEGKW-LVLGEDAAVTG-SQTING-----QHLVF--1248
QY 434 VEQANWDFNLTGSGVVEKQKASIRYDKALIDRENYKQPLDLCAHROTARSAFA--LSG 491
DB 1249 --RAN-----GVQVKGEFVTDYR-----GRISY-----DSNSGQIRNRFVRNAQG 1288
QY 492 NMVYFTPOHKLSTASHQERLPSTQELAHGKHVATNTFEVNGKHLNRSNNIELALGYE 551
DB 1289 QMFYFDNNGYAVTGA---RTINGQHLFRANGVQVKGEFVTDH-----GRISYDGN 1339
QY 552 GDBWQYNLALYRNP-----GNVYIAQLNDRGPKSTEDDSEMKLVRYNOSGADFYGA 606
DB 1340 GDQ-----IRNRFVRNAQGWQFYFD--NNG-----YAVTGARTINGQ 1374
QY 607 GEIYFKPTPRYRIGYSGDYVR---GRKLNLPGLPREDAYGNRPFI--AQD-----DONA 656
DB 1375 -HLYFEAN---GVQVKGEFVTDYRGRISYDNSG--DQIRNR-FVRNAQGWQFYFDNNG 1427
QY 657 PRVPAARL--GPHL--KAG-----LTDRIDANLDYY 683
DB 1428 YAVTGARTINGQHLFRANGVQVKGEFVTDYR-GRISYY 1465
RESULT 32
ID LBPA_NEIMA STANDARD; PRT; 944 AA.
AC AC
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin binding protein A precursor.
GN LBPA OR NMA1739.
OS *Neisseria meningitidis* (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; *Neisseria*.
OC NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,


```

QY 536 LNKERSNNRIEALGVGDRWQYNLALYRNRFGNYIAQTLDNDRGPKSIEDDSEMKLVRY 595
Db 473 -----ELVVGTSAS-----FSHW-----EGKSYWNLRY 496
QY 596 NOSGADFYGAEGEIVFKPTPRYRIGVSGDYVRGRILNPLSLPGRDAYGNRPFFIAQDDQN 655
Db 497 DNTTDDFINWGDIL-GKPD-----WGTPSQYIDDKTRQLGS-----YWTARFNVTTDD-- 542
QY 656 APRVPAARGLHKLASLTDRIANLDYRVRFAQNKLARYETRTPGHMLNLGANYRNRTR 715
Db 543 -----LNLFLGGRVVD-----YRVTGLNP-----TIRESGRFPIYGVAVVDLNDT 582
QY 716 YGEW-----NWYVADNLLNQSVYAHSSFLSTPQMGSRFTGGVNVKF 758
Db 583 YSVYASYTDIFMPQDSWYRDSNKLLE-----PDEGQYIEIGIKGEY 624

```

RESULT 34

```

HTPG_SALTY
ID HTPG_SALTY STANDARD; PRT; 624 AA.
AC P59480;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chapterone protein htpG (Heat shock protein htpG) (High temperature
protein G).
GN HTPG OR STM0487.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]

```

```

SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RA MEDLINE=21534948; PubMed=11677609;
RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston K., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856 (2001).
CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL; AE008718; AAL19441.1; ALT_INIT.
DR StyGene; SG????; htpG.
DR HAMAP; MF_00505; -; 1.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR001404; Hsp90.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00183; HSP90_2.
DR PRINTS; PR00775; HEATSHOCK90.
DR SMART; SM00387; HATPase_c; 1.
DR PROSITE; PS00298; HSP90; 1.
KW Chaperone; ATP-binding; Heat shock; Complete proteome.
FT DOMAIN 1 336 A; SUBSTRATE-BINDING (BY SIMILARITY).
FT DOMAIN 337 552 B (BY SIMILARITY).
FT DOMAIN 553 624 C.
SQ SEQUENCE 624 AA; 71486 MW; 0974B2B1B38D0B39 CRC64;

```

```

Query Match 3.1%; Score 123.5; DB 1; Length 624;
Best Local Similarity 20.1%; Pred. No. 0.58;
Matches 150; Conservative 85; Mismatches 293; Indels 219; Gaps 32;

QY 54 STASDKIISGDTLRQKAVNLGDAL--DGVPGIHASQYGGGASAPVIRGTG-RRKVLNLH 110
Db 37 SNASD---AALKRFRALSNFPLYBGGELRVRSFDDKRTLTITADNGVGNRDEVIDH 93
QY 111 HGETGDMADFDPDIAIMVDTALSSQVEILRGFVTL-LYSSGNVAGLVADVADGKIPKQPE 169
Db 94 LGTTAKSTKTFLESMSGDQAKDSQ---LIQFGVGFSAFIVADKVTVRRAADGK-PE 149
QY 170 NGVSGELGRLSSGNLEKLTSGGINIGLGNKFNVLHTEGLYKRSQDYAVPRYRNLKRLPDS 229
Db 150 NGVFEW-----SAG-----EGEYTVADITKDRGTETI 176
QY 230 PRRFANGQHRAVLGWRKFPYRRTYSDDRDQVGLPA--HSHEYDDCHADIIWQKSLINK-- 285
Db 177 TLHLRGEDEFLDDMRVRSIIISKYS---DHIALPVEIEKREKDGGETVISWEK--INKAQ 231
QY 286 -----RYLOLYPHLLTBEEDVDYDNPGLSCGFHDDDDDAHAHAHNGKPWIDLRN 332
Db 232 ALWTKNKSEIKDDEYNFVKHIA-----HDFDTPLTWSHN----- 266
QY 333 KRYELRAEWKQFPFGFEALRVHLNRNDYHDEKAG-----DAVENPFNNQOTQNA 381
Db 267 -----RVEGKQETSLLYIPSAQPDWDLNRDHGHLKLYQVRVFTMDDAEQFMPNYLRFV 321
QY 382 R--IELRHQPIG-----RLKSGWGVQYLCQKSALSATSEAVKQPMLLDNKVQHYHFFGVE 435
Db 322 RGLIDSNPLNVSREILQDSTVTNRNL---RSALTQRYVLMLEKLAKODAEKYQTF---- 374
QY 436 QANWDNFTLEGVRVEKQKASIRYDKALIDRENYKQPLDPLGAHRQTARSPALSGNMYF 495
Db 375 ---WKQF---GLVLKEGPAEDHANQEAIAKLLPASTHTDSSAQTVSLEDY----- 419
QY 496 TPQHKLSUTASHQERL-PSTOELYAHGKHVATNTFEVGNKHLNKNERSNNIE-LALGYEGD 553
Db 420 -----VSRMEGQEKIYIITADSYAAAKN-----SPHLELRKKKGIEVLLLSDRID 465
QY 554 RWQYNLALYRNRFGNYIAQTLDNDRGPKSI--EDDSEMKLVRYNQSGADFVGARGEIYF 611
Db 466 EWMN---YLTEF-----DGKAFQSAKADESIEKLA-----DEVENAKAEKAL 508
QY 612 KPTPRYRIGVSGDYVRG-----RLKNPLSLPGRDAYGNRPPIAQDDQNAQPRVPAARLGF 666
Db 509 EPFVERVKTLTGDRVKDVRVLRHLTDTPAI----- 538
QY 667 HLKASLTDRIANLDYRVRFAQNKLARYETR-----TPGHMLNLGANYRNRTRYGEWNW 721
Db 539 -----VTTDAEMSTQMAKLFAAAGQSVPEVKYIFELNPDHVLVVKRTADTKDQAKQEW-- 592
QY 722 YVKADNLLNQSVYAHSSFLSTPQMG 748
Db 593 ---VELLDQALFAERGTLDPNQFIR 616

RESULT 35
HTPG_SALTY
ID HTPG_SALTY STANDARD; PRT; 624 AA.
AC P59479;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chapterone protein htpG (Heat shock protein htpG) (High temperature
protein G).
GN HTPG OR SFY0531 OR T2373.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.

```



```

QY 332 N---KRYELRAEWKQPPGPEALRVHLNRNDYHDEKAGDAVENFFNNQTONARIE-LR 386
DB 215 NGIYKNUHDL-----ADLNHNSTSDVYLKDKAKWLDLIGDIGIRMDAVK 259
QY 387 HOPICRLK-----GSW--GVQYLGQKSSALSATSBAVKQPMLLDNKVQHY 429
DB 260 HMPFGQKSFMAAVNNYKPVFTFGWFLGVNEVGPENHKFANESGMSLLDPRFAQVRQ- 318
QY 430 SFFGVEQANWDF-----TLEGG-----VRVEKQASI-RYDKALIDRENNYKQPIPDIGA 479
DB 319 ----VFRDNTDMYGLKAMLEGAADYQVDDQVTFIDNHDMERFHASNANRKL----- 369
QY 480 HRQTARSPAL-----SGNNYFTPOHKLST-ASHOERLPSTQELVYAHGKHVATNTEFVGN 533
DB 370 --EQLAFLLILARPAIYGTQYMSGGTDPDRARIPFS-----TSTTAYQVIQ 419
QY 534 KHLNKRNNIBELAGYEDRQYNN-LALYRNRFGNIYAQTLNDG-RGPKSIED-DSEM 590
DB 420 K-LAPLRKSNPAIAYGSTQERINNDVLIYERKFGSNVAVVNRNLNAPASISGLVTSL 478
QY 591 KLVRYNQ----- 597
DB 479 POGSYNDVLGGLLNGNTTVSGGGAASNFTLAAGTAVQVYTAATPTTIGHVGPMAKP 538
QY 598 ----SGADFYGARGEIYFKPTPRYRIGVSGDYVRGRKLNPLSLPGREDAYGNRPFFIAQ 651
DB 539 GVTIITIDRGFGSSKGTYYFGTT-----AVSG-----ANITSW-----E 572
QY 652 DDQNAAPVPAARLG-FHLKASLUTRIDANL-DYRVVFAQNKLA-----RYEFTPTGHEML 704
DB 573 DTQIKVKIPAVAGGIYNIKVANAAGTASNVDYDFEVLSDGVSRVFNNAATLQGLY 632
QY 705 NLGAVYRNTRYGWNVYVKADNLNQSVAHSSFLSDTPQMGSRFTGQVNVKP 758
DB 633 LTG-----NVSELGNWDPAKAGPMNQVYQIPNMYDV-----SVPAKGTIEF 677

```

RESULT 38

```

ID YDDB ECOLI STANDARD; PRT; 790 AA.
AC P31827; P77572;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yddb (CDS103).
GN YDDB OR B1495.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12.
RA Tulin E., Gasser F., Biville F.;
RT "Sequence and functional analysis of an Escherichia coli DNA fragment
RT able to complement pqeE and pqgF from Methylobacterium organophilum.";
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,

```

```

RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasundaram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377 (1996).
CC -!- SIMILARITY: TO H.INFLUENZAE HT1369.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, X71917; CAA50733.1; -.
DR EMBL; AE000246; AAC74568.1; -.
DR EMBL; D90791; BAA15166.1; -.
DR PIR; B64903; B64903.
DR EcoGene; EG11743; yddb.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rc; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 1 22 MKRVLPGLVILGADVAQAVDD -> MT (IN REF. 1).
FT CONFLICT 279 289 TWFSADRTHD -> YLVCORPLYPR (IN REF. 1).
FT CONFLICT 503 503 A -> T (IN REF. 1).
FT CONFLICT 534 545 TLTRYQDLKTPY -> NSDALSGFE (IN REF. 1).
FT CONFLICT 573 574 HD -> LV (IN REF. 1).
FT CONFLICT 615 616 HI -> LY (IN REF. 1).
SQ SEQUENCE 790 AA; 89282 MW; 907735A2A124E674 CRC64;
Query Match 3.0%; Score 120.5; DB 1; Length 790;
Best Local Similarity 19.8%; Pred No. 1.3; Mismatches 104; Indels 271; Gaps 47;
Matches 175; Conservative 104;
QY 16 INTPLLAQAHEQTSVGLTETVTVGKSRPRAT---SGLLHTSTASDKIISGTLRQAVN 72
DB 39 VVPVNGNTHYTSESI-----ERLPTGNGNISDLLRTNPA---VRMDSTQSTSLN 85
QY 73 LGDALDGVPGIH-ASQYGGGASAPVIRGTGRIRKVLNHHGTGMDAFSP-----DH 124
DB 86 QGDIRPEKISIHGASPY---QNAVILDG-----ISATNNLNPNPE-SDASSATNIGMSQ 136
QY 125 AIMVDLTALSQQVEILRGVPVTLTYSSGNVAGVDVA-----DGKI-----PE 165
DB 137 GYLLDVSLDNTLYDSFVPEFGEFN-GVIDAKIKFPNADDKVKLGRTTRSDWLTS 195
QY 166 KMPENGVS---GELGLSLSSGNLEK-LTSGGINIGLGNFVLHTEGLYRKSQDYAVPRY 220
DB 196 HIDDENKSAFNQSGSGSTYSPDFKKNFYTLISFQELADNFGV-TAGLSRRQSDITRADY 254
QY 221 -----RNLKELPD-----SPRFANGQHRVILGWRKFRVRY--SDR--- 256
DB 255 VSDGIVAGRAQYKNVIDTALSKEFTWFASDRFTHDLTLKYTGSSRDYNTSTFPQSDREMG 314
QY 257 RDQGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTBEDVDYDNPGLSCGFHDDDD 316
DB 315 NKSYGLAW-----DMDTQLAWAKLRTTVGWDHSDYTRHDDHWYTE--LSCITYGDI 366
QY 317 AHAHANGKPIDLRNRYELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFFNN 376
DB 367 RCTRGGLGHISQAVDNTVFKRLDW-QKF-----AVGN----- 398
QY 377 QTONARIELRHOP-IGR---LKGSGVOYLGQKSSALSATSBAVKQPMLLDNK----VOH 428
DB 399 -----VSHQPYFGAEIYSDAWTERHNSQSESVINAAKTNHTIYHKGKRLGIDN 450
QY 429 YSFFGVEQANWDFTELEGGVREKQASIRYDKALIDRENNYKQPIPDIGHAHTARSFA 488
DB 451 YTLVMDRISWRNVSLMPGV-----YDYN-----LSNHNISPRFMT 489

```

QY 489 LSGNWTTPQHKLSLPSHQRLEPSTQELYAHGKHVATNFEV-----GNKHLNK- 538
Db 490 ---EMDIFANQSMITAGYNR-----YGGNILDGMGLDIRNSWTSVSGNKLTRY 538
QY 539 ---ERSNNELALGYEGDRWOYNLALYRNFNGYIIAQTLNDGRGPKSIEDSEMK--LV 593
Db 539 QDLKTPYNDELAWGLOOKIGKNVIAR-----ANYVIREA--HDQISKSRTSDATKTTIT 591
QY 594 RYNQSG-----ADFYGAEGEIIYFKPTPRYRIGVSGDYVRGLKNLPSL----- 636
Db 592 EYNNDGKTKTHSPSLSELAEE-----PLHIRQVDINPOLIVFSYIKSGKNLSLNNGYEES 645
QY 637 -PQREDAYGNRPFIADODQNAQPRVPAARLGFHLKASLTDRIANLDYR-----VFAQNK 690
Db 646 NTGDNQVYNGNLVSYDS-----VPVADFNNPLKISL-----NMDFTHQPSGLVWA-NT 693
QY 691 LARYETRTGCHMLNLG-ANYRNRTRYGWNWTVKADNLLNQSIVYAHSSFLSDTPQM--- 746
Db 694 LAWEARKA---RIILGKTNAQVISEYSDYKQYV--DEKLDSSL-TWDTLSWTPOFLOQ 747
QY 747 -----GRSFTGGVNVKF 758
Db 748 QNLITISADILNVLDSKTAVDTTNTGVATYASGRFTWLDVSMKF 790

RESULT 39
GELS_HORSE
ID GELS_HORSE STANDARD; PRT; 730 AA.
AC Q28372; Q95180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Gelsolin (Actin-depolymerizing factor) (ADF) (Brevin).
GN GSN.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Smooth muscle;
RX MEDLINE=98149677; PubMed=9490033;
RA Koepf E.K., Hewitt J., Vo H., Macgillivray R.T.A., Burtnick L.D.;
RT "Equus caballus gelsolin -- cDNA sequence and protein structural
RT implications.";
RL Eur. J. Biochem. 251:613-621(1998).
CC -!- FUNCTION: Calcium-regulated, actin-modulating protein that binds
CC to the plus (or barbed) ends of actin monomers or filaments,
CC preventing monomer exchange (end-blocking or capping). It can
CC promote the assembly of monomers into filaments (nucleation) as
CC well as sever filaments already formed.
CC -!- SUBUNIT: BINDS TO ACTIN AND TO FIBRONECTIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE VILLIN/GELSOLIN FAMILY.
CC -!- SIMILARITY: Contains 6 gelsolin-like repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U31699; AAC13353.1; -.
DR FDB; 15-SEP-99.
DR InterPro; IPR001974; Gelsolin.
DR Pfam; PF00626; Gelsolin; 6.
DR PRINTS; PR00597; GELSOLIN.
DR SMART; SM00262; GEL; 6.
KW Cytoskeleton; Actin-binding; Repeat; Calcium; Capping protein;
KW 3D-structure.

FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 1 124 ACTIN-BEAVING (POTENTIAL).
FT REPEAT 382 730 ACTIN-BINDING, CA-SENSITIVE (POTENTIAL).
FT REPEAT 24 74 GELSOLIN-LIKE 1.
FT REPEAT 146 186 GELSOLIN-LIKE 2.
FT REPEAT 262 304 GELSOLIN-LIKE 3.
FT REPEAT 401 452 GELSOLIN-LIKE 4.
FT REPEAT 524 564 GELSOLIN-LIKE 5.
FT REPEAT 627 669 GELSOLIN-LIKE 6.
FT SITE 71 74 ACTIN-ACTIN INTERFILAMENT CONTACT POINT.
FT SITE 110 117 POLYPHOSPHONOSITIDE BINDING
FT SITE 136 144 POLYPHOSPHONOSITIDE BINDING
FT SITE 136 144 POLYPHOSPHONOSITIDE BINDING
SQ SEQUENCE 730 AA; 80696 MW; F5479F2235046A4D CRC64;
Query Match 3.0%; Score 119.5; DB 1; Length 730;
Best Local Similarity 21.3%; Pred. No. 1.4; Indels 117; Gaps 20;
Matches 96; Conservative 61; Mismatches 177;
QY 322 HNGKPMI-DLRNRYELRAEWKQPPFGPEALRVHLNRNDYHHDEKAGDAVENFFNQTON 380
Db 159 NNGDCFILDLGNIIYQWCGSKSNRPERLKATOVSKGIRD---NERSGRAQVSVFERGAEP 215
QY 381 ARIELRHQPIGLKSGWGVYLGOKSSALSATSEAVKQ-----PMLL 422
Db 216 EAM-----LQVLGPKPTLPEATEDTVKDAANRKLAKLYKVSNGAGPMV 260
QY 423 DNKQVQHSYFQVQANWDMNFTLEGG-----VRVEKOKASIRYDKALIDRENY 470
Db 261 SLVADENPPAQGALSREDCFILDHGKDGKIFWKQKQANMERKAAKKTASDFISKMDVP 320
QY 471 KQP-----LPDLGAHQRTARSFALSGNWTPTQHK-----LSLTASHQERLP 512
Db 321 KQTQSVLPE-GGETPLPRQF--FKNWRDPDQTEGLGLAYLSSHIAHVERVPDAATLHT 377
QY 513 STQELYAHGKHVATNFEVGNKHL-NKERSNNIEL-----ALCYEGDRVQYNALYRNP 566
Db 378 STAMAAQHG-----MDDGTGQKQIWRVGSNKVPDVPATYQGYGDSY---IILYNYRH 430
QY 567 GNYIYAQTLNDGRGPKSIEDSEMKLVRYNQSGADFYGAEGEIIYFKPTPRYRIGVSGDYV 626
Db 431 GSR-QGQIIYNWQGAQSTQDEVAASAILTAQDDELGG-----TP-----V 470
QY 627 RGRLNKLPSPGRED-----YGNRPFI-----AQDDQNAQPRVPAARLGFHLKASLTDRID 677
Db 471 QSRV-----VQGEPAHLMSLFGKPMIYVKGTSREGQTAPASTRLFQVRASSSGATR 525
QY 678 ANLDYRVYFAQNKLYAREYTRTPGHMLNLGA 708
Db 526 AVEIIPKAGALNSNDAFVLKTPSAAYLWVGA 556

RESULT 40
YC17_HAEIN
ID YC17_HAEIN STANDARD; PRT; 913 AA.
AC P45114;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable tonB-dependent receptor H11217 precursor.
GN H11217
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness B.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

Result No.	Score	Query Match	Length	DB ID	Description
1	3845	95.3	764	16 Q9JUS2	Q9JUS2 neisseria m
2	3836	95.0	758	16 Q9JZN9	Q9JZN9 neisseria m
3	1724	42.7	766	2 Q9AER3	Q9AER3 pasteurella
4	1492.5	37.0	809	16 Q9C1W8	Q9C1W8 pasteurella
5	1364	33.8	805	16 Q9C120	Q9C120 pasteurella
6	1307	32.4	925	16 Q9CMR7	Q9CMR7 pasteurella
7	745	18.5	688	16 Q8P7F3	Q8P7F3 xanthomonas
8	718.5	17.8	687	16 Q8PIR6	Q8PIR6 xanthomonas
9	694.5	17.2	681	16 Q9PGB8	Q9PGB8 xylella fas
10	679	16.8	687	16 Q915F7	Q915F7 pseudomonas
11	636	15.8	687	16 Q9ABL3	Q9ABL3 caulobacter
12	395.5	9.8	777	16 Q8P1I6	Q8P1I6 leptospira
13	354.5	8.8	778	16 Q9CLF3	Q9CLF3 pasteurella
14	290	7.2	713	16 Q9KNA1	Q9KNA1 vibrio chol
15	264.5	6.6	706	2 Q93TP3	Q93TP3 escherichia
16	264.5	6.6	715	16 Q8FGA0	Q8FGA0 escherichia

```
Db 7 MAQTTLKPIVLSILLINTPLLSQAHGTEQSGLETVSVVGKSRPRATSGLLHTSTASDKI 66
QY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTGRRIKVLNHHGSETGDMADF 120
Db 67 ISGDTLRQKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTGRRIKVLNHHGSETGDMADF 126
QY 121 SPDHAIMVDTALSQQVEILRGPTVLLYSNGVAGLVDVADGKIPEKMPENGVSGLGLRL 180
Db 127 SPDHAIMVDSALSQQVEILRGPTVLLYSNGVAGLVDVADGKIPEKMPENGVSGLGLRL 186
QY 181 SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSQYAVPRYRNKLRLPSRPFANGQHRA 240
Db 187 SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSQYAVPRYRNKLRLPS---HADSQTGS 243
QY 241 V-LGW--RKRFRYRTYSDRRQYGLPAHSHEYDDCHADIWQKSLINKRYLOLYPHLLTE 297
Db 244 IGLSWVGKGFIGAAYSDRRQYGLPAHSHEYDDCHADIWQKSLINKRYLOLYPHLLTE 303
QY 298 EDVDYDNPGLSCGCFHDDDDAHAAHNGKWPIDLRNKRVELRAEWKQPPGPFALRVHLNR 357
Db 304 EDIDYDNPGLSCGCFHDDDDAHAAHNGKWPIDLRNKRVELRAEWKQPPGPFALRVHLNR 363
QY 358 NDYHDEKAGDAVENFNNQTNARIELRHQPIGRLKSGWQYVYLGQKSSALSATSEAVK 417
Db 364 NDYHDEKAGDAVENFNNQTNARIELRHQPIGRLKSGWQYVYLGQKSSALSATSEAVK 423
QY 418 QPMLLDNKVQHYSPFGVEQANWDFLEGVVRVEKQASIRYDKALIDRENYVKOPLDPL 477
Db 424 QPMLLDNKVQHYSPFGVEQANWDFLEGVVRVEKQASIRYDKALIDRENYVKOPLDPL 483
QY 478 GAHQRTARSFALSGNWYFTTPQHKLSLTASHOERLPSTQELYAHGKHVATNTPEVGNKHLN 537
Db 484 GAHQRTARSFALSGNWYFTTPQHKLSLTASHOERLPSTQELYAHGKHVATNTPEVGNKHLN 543
QY 538 KERSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSMKLVRYNQ 597
Db 544 KERSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSMKLVRYNQ 603
QY 598 SGADFYGAEGEYIFPKTPRYRIGVSGDVYRGKLNPLSLPGREDAYGNRPFFAODQONAP 657
Db 604 SGADFYGAEGEYIFPKTPRYRIGVSGDVYRGKLNPLSLPGREDAYGNRPFFAODQONAP 663
QY 658 RVPAAALGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGANRYNTRYG 717
Db 664 RVPAAALGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGANRYNTRYG 723
QY 718 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 724 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 764
```

RESULT 2

```
Q9JZN9 PRELIMINARY; PRT; 758 AA.
ID Q9JZN9
AC Q9JZN9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE TonB-dependent receptor.
GN NMB0964
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
NC NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittiome H., Clark E.B.,
```

RESULT 3
Q9AER9

```
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiognani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002448; AAF62323.1; -.
DR TIGR; NMB0964; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 758 AA; 85119 MW; 6BA342986968640B CRC64;
```

Query Match Score 3836; DB 16; Length 758;

Best Local Similarity 95.7%; Pred. No. 3.1e-268;

Matches 728; Conservative 9; Mismatches 18; Indels 6; Gaps 3;

```
QY 1 MAQTTLKPIVLSILLINTPLLSQAHGTEQSGLETVSVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MAQTTLKPIVLSILLINTPLLSQAHGTEQSGLETVSVVGKSRPRATSGLLHTSTASDKI 60
```

```
QY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTGRRIKVLNHHGSETGDMADF 120
```

```
Db 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGGASAPVIRGQTGRRIKVLNHHGSETGDMADF 120
```

```
QY 121 SPDHAIMVDTALSQQVEILRGPTVLLYSNGVAGLVDVADGKIPEKMPENGVSGLGLRL 180
```

```
Db 121 SPDHAIMVDTALSQQVEILRGPTVLLYSNGVAGLVDVADGKIPEKMPENGVSGLGLRL 180
```

```
QY 181 SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSQYAVPRYRNKLRLPSRPFANGQHRA 240
```

```
Db 181 SSGNLEKLTSGGINIGLGNFVLHTEGLYRKSQYAVPRYRNKLRLPS---HADSQTGS 237
```

```
QY 241 V-LGW--RKRFRYRTYSDRRQYGLPAHSHEYDDCHADIWQKSLINKRYLOLYPHLLTE 297
```

```
Db 238 IGLSWVGKGFIGAAYSDRRQYGLPAHSHEYDDCHADIWQKSLINKRYLOLYPHLLTE 297
```

```
QY 298 EDVDYDNPGLSCGCFHDDDDAHAAHNGKWPIDLRNKRVELRAEWKQPPGPFALRVHLNR 357
```

```
Db 298 EDIDYDNPGLSCGCFHDDDDAHAAHNGKWPIDLRNKRVELRAEWKQPPGPFALRVHLNR 357
```

```
QY 358 NDYHDEKAGDAVENFNNQTNARIELRHQPIGRLKSGWQYVYLGQKSSALSATSEAVK 417
```

```
Db 358 NDYHDEKAGDAVENFNNQTNARIELRHQPIGRLKSGWQYVYLGQKSSALSATSEAVK 417
```

```
QY 418 QPMLLDNKVQHYSPFGVEQANWDFLEGVVRVEKQASIRYDKALIDRENYVKOPLDPL 477
```

```
Db 418 QPMLLDNKVQHYSPFGVEQANWDFLEGVVRVEKQASIRYDKALIDRENYVKOPLDPL 477
```

```
QY 478 GAHQRTARSFALSGNWYFTTPQHKLSLTASHOERLPSTQELYAHGKHVATNTPEVGNKHLN 537
```

```
Db 478 GAHQRTARSFALSGNWYFTTPQHKLSLTASHOERLPSTQELYAHGKHVATNTPEVGNKHLN 537
```

```
QY 538 KERSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSMKLVRYNQ 597
```

```
Db 538 KERSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSMKLVRYNQ 597
```

```
QY 598 SGADFYGAEGEYIFPKTPRYRIGVSGDVYRGKLNPLSLPGREDAYGNRPFFAODQONAP 657
```

```
Db 598 SGADFYGAEGEYIFPKTPRYRIGVSGDVYRGKLNPLSLPGREDAYGNRPFFAODQONAP 657
```

```
QY 658 RVPAAALGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGANRYNTRYG 717
```

```
Db 658 RVPAAALGFHLKASLTDRIDANLDYRVFAQNKLARVETRTPGHHMLNGANRYNTRYG 717
```

```
QY 718 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
```

```
Db 718 EWNVYKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
```

Q9AER9	PRELIMINARY;	PRT;	766 AA.	
Q9AER9	Q9AER9	Q9AER9	Q9AER9	Q9AER9
DT	01-JUN-2001 (TREMELrel. 17, Created)			
DT	01-JUN-2001 (TREMELrel. 17, Last sequence update)			
DT	01-MAR-2003 (TREMELrel. 23, Last annotation update)			
DE	Iron-regulated outer membrane protein.			
DE	IRP.			
GN	Pasteurella haemolytica.			
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Mannheimia.			
OX	NCBI_TaxID=75985;			
RN	SEQUENCE FROM N.A.			
RP	InterPro; IPR000531; TonB_boxC.			
RA	Lo R.Y., Graham M.R.;			
RT	"Putative TonB dependent receptor of Mannheimia haemolytica.";			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DDB databases.			
CC	-!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).			
DR	EMBL; AY028475; AAK29743.1; -			
DR	InterPro; IPR000531; TonB_boxC.			
DR	Pfam; PF00593; TonB_dep_Rec; 1.			
DR	PROSITE; PS00430; TONB DEPENDENT REC 1; 1.			
KW	Membrane; Outer membrane; Receptor; TonB box.			
SQ	SEQUENCE 766 AA; 86440 MW; 7D76FAE05D7C5619 CRC64;			
Query Match				
Best Local Similarity 44.6%; Pred. No. 1.le-115;				
Matches 346; Conservative 128; Mismatches 260; Indels 42; Gaps 12;				
QY	9	IVLSILLINTPLLAQAHETEQSGVLETVVVGKSPRATSGLLHTS-TASOKIISGDTLR	67	
Db	7	LTVSILLIISG-VGVANE-----ISLETTVDGNT-PSYKGLGGLNSNESVDEKMLK	60	
QY	68	QKAVNLGDALGVPGHAGSYGGASAPVIRGQTRIKVNLHGETGDMADFPDPAIM	127	
Db	61	QGSITLGNALSGELGHSHSQFGGASTPIRQGESKRAKILQNGENLDMGMSPDHAT	120	
QY	128	VITALSQVEILRGPTVLLSYSSNAGLVADGKIPKMPENGVSGLGLRLSSGNLEK	187	
Db	121	VDALLAKLEILRGPTVLLSYAGTAGVINVDNKIPAIPEKGYEGGQVRFSGASKE	180	
QY	188	LTSGGINIGLKNFVLHTEGLYKSGDYAVPRY-----RNKRLPDSPPRFANGQRAVLG	243	
Db	181	LTVAGSTFALGNHALRVQGMYNKASEYVAPHTIEGKPYHRVPSDVQSQGT--TVSL	238	
QY	244	W--RRFYRTYSDRRDQGLPAHSHYDDCHADIIWQSLINKYQLIYPHLLTEEDVD	301	
Db	239	WIGERHGLGIATYDRDKYGLTGHTKDYHTISIRQAWMPAKGYLRFPYPLABEGD	298	
QY	302	YONPGLSC-----GFHDDDAHAHNGKPKWIDLNRKRYELRAEWKQPPFGFEALRVH	354	
Db	299	YANPGIRLLAHTIPGSGHYQDTHF--GKPIDMHSKRYDIDGSLQNLPLFGFEAKIS	355	
QY	355	LNENDYHDEKAGDAVENFNQTONARIELRHQPIGLKSGWGYVYLGOKSSALSATSE	414	
Db	356	ANYVDYHDEKQKRYENTFKNGKRLRFLVHKWKGLKGAIGVQVYTNQSTALALEAS	415	
QY	415	AV-----KQPLLDNKHQVSYFPGVQANWNTFTLEGVVRVEKQKASIRYDKALIDRENY	469	
Db	416	RAAKVFNKQPLNNPKTKLWSLPAIELNLGDTFELSGAERQKIAMDYDKLIDRWLG	475	
QY	470	YKQPLDLGAHQRTASFGALSGNWYTPPHQKSLTASHQERLPSTQELYAHQKHVATNTP	529	
Db	476	FNTPMNLPHDKGYSYFATHWYFAPNHKLTUNAHHQERLPNAQELYAHQKHIALNAF	535	
QY	530	EVGNKHLNERSNNIELALGYEDRWQYNLALVRNFGNVIYAQTLDNGRGPSTEDDSE	589	
Db	536	EAGNKLKERSNQIELSLIAYGDKWDYKLNLTHTHTYIGNYIPLTLDNRGPKSTDEYN	595	
QY	590	MKLVRYNQSGADFYAGBEIYFXTPTRYRIGVSGDYVRGKLNKLPSTL-----PQREDA	642	
Db	596	LKNRYIYQGEARPSGAEGEIGLYFTPNYRLAVPGDVVRGKLNLPNIAMSYNIWTGEVDK	655	
QY	643	YGNRPFAQDDQNAVRPAARLGFHLKASLTDRIDANLIDYRVRFAQKRLARTETTPGH	702	
QY	656	WASQP-----DISAPRIPPLRLGARNADFNLWNSGMLERYRFAQKRVSKYEQVTPGH	710	
QY	703	MLNLGANYRNTYRGWNVVYKADNLNOSVVAHSSFLSDTPQMGSRFTGGVNVKF	758	
Db	711	QVNLGVTSNNHFNQTEYQVFLKVDNLNKLQYHASYLPHIPQMGGRNAMLGNWISF	766	
RESULT 4				
Q9CLW8	PRELIMINARY;	PRT;	809 AA.	
AC	Q9CLW8	Q9CLW8	Q9CLW8	Q9CLW8
DT	01-JUN-2001 (TREMELrel. 17, Created)			
DT	01-JUN-2001 (TREMELrel. 17, Last sequence update)			
DT	01-MAR-2003 (TREMELrel. 23, Last annotation update)			
DE	Hypothetical protein PM1081.			
GN	PM1081.			
OS	Pasteurella multocida.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Pasteurella.			
OX	NCBI_TaxID=747;			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=PM70.			
RC	MEDLINE=21145866; PubMed=11248100;			
RA	May B.J., Zhang Q., Li L.B., Paustian M.L., Whittam T.S., Kapur V.;			
RL	"Complete genomic sequence of Pasteurella multocida PM70.";			
DR	Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).			
DR	EMBL; AE006149; AAK03165.1; -			
DR	InterPro; IPR000531; TonB_boxC.			
DR	Pfam; PF00593; TonB_dep_Rec; 1.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 809 AA; 90903 MW; 23028B7DFCBF14DB CRC64;			
Query Match				
Best Local Similarity 37.0%; Score 1492.5; DB 16; Length 809;				
Matches 333; Conservative 120; Mismatches 271; Indels 53; Gaps 13;				
QY	27	TEQSVGLETVTVGKSPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDVPFGHAS	86	
Db	41	TSEHIEIFVNTLIBESREGAPLGRLMASEKIIIPAYSKQSGNLGDSLSSELGHAS	100	
QY	87	QYGGASAPVIRGQTRIKVNLHGETGDMADFPDPAIMVDTALSOQVEILRGPTVLL	146	
Db	101	QYGGASAPVIRGQTRIKVNLHGETGDMADFPDPAIMVDTALSOQVEILRGPTVLL	160	
QY	147	YSSGNVAGLVADGKIPKMPENGVSGLGLRLSSGNLEKLTSGGINIGLKNFVLHTE	206	
Db	161	YSSGNAAGVNVVDNKIP-TAEVVVGVEGVLRTGSADNERLVNVALDVLGSKHFALHLE	219	
QY	207	GLYKSGDYAVPRY---NLKRLPDSPPRFANGQRAVLG--RKFYRTYSDRRDQ	259	
Db	220	GLHKKAGDYRTPSYQSGSTHKLANS---FVDRSGVGLSWGDKGYLGVAYSQRDK	276	
QY	260	YGLPAHSHYDDCHADIIWQSLINKYQLIYPHLLTEEDVDVNDPGLSC-----GFHDD	314	
Db	277	YGLPAHSHLYDEYVMVLLSDAHRKPYLKHYPFMEETDIDYNNPNDICIKKEMSHSHG	336	
QY	315	DDAHAAHNGK-----PWIDLRNKRVELRAEWKQPPFGFALRVHLNRNDYH	362	
Db	337	LCNHGAHNGQSHDHHADPHIALNTQRMDLRGWKNPKVGLDKVFSKAVGYRH	396	
QY	363	DEKAGDAVENFNQTONARIELRHQPIGLKSGWGYVYLGOKSSAL-SATSEAVKQPL	421	
Db	397	DEKSGALSDNSFKNGYSARVEFLHQPIAGVGLIGLSHYQDSYALDMHTLEYKQNL	456	
QY	422	LDNKVQHSYFPGVQANWNTFTLEGVVRVEKQKASIRYDKALIDRENYKQPLDL-GAH	480	
Db	457	SDHTTAQCSLFMEHVELGKMQFDIGRVEKQRIAMKYHFNVPKDE---QPPEELTRPH	512	
QY	481	QRTARSFALSGNWYFTPHKLSLTASHQERLPSTQELYAHQKHVATNTEFGNKHAKER	540	
Db	513	KSAYSALSNYQNLNQHFNMIVSHQERLPNAQELYAHQKHIALNFSFAGNKNLYKER	572	


```

QY 541 SNNIELALGVEGRWYNNIALYRNFRGNYIAOTLN-----DGRGPKSIEDSEMK 591
Db 573 SNNVELGWGTGKGLGKLSGGYQQFSNYIYAAILNNKTCSPWRPNRSLRSLSDYPLR 632
QY 592 LVYRNSGADFYGAEGEIVFKPFRVIRGVSGDYVRGLKNLPSP-----GREDAY 643
Db 633 LYRYNQAKIYGLEAEVSQISSTHSVIFG DYVRGKLKDLPSLPIGYKYIYENYDMV 692
QY 644 GNRP--FIAQDDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETTPGH 701
Db 693 GVQPTGWEKQPDGNAPRMSPMRLGIKWNAYFDNGISFNTQLXRVFAQNKVARLETPTKG 752
QY 702 HMLNLGANYRNRTRYGEMWVYKADNLNQSVVAHSSFLSDTPQMGRSFTGGVNVKPF 758
Db 753 TMLNLGMSYDGKMGNNETTLFANVNVNLSRVNHTSFLSYIFQSGLGLNVGMNFKF 809

RESULT 5
Q9CL20 PRELIMINARY; PRT; 805 AA.
AC Q9CL20;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PM1428.
GN PM1428.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006179; AAK03512.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep Rec; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 805 AA; 90850 MW; 799A05B9E4D10509 CRC64;

Query Match 33.8%; Score 1364; DB 16; Length 805;
Best Local Similarity 37.3%; Pred. No. 1.3e-89;
Matches 294; Conservative 136; Mismatches 291; Indels 68; Gaps 10;

QY 26 ETEQSGVLETVTVGKSRPRATSGLLHTSTASDKTIISGDTLRQKAVNLGDALDGVPGIHA 85
Db 29 EKKQIVFDEISVESKGAFRSDPLSGLPKQNDILVSKQKLKTSSTLGNALAGELSVHS 88
QY 86 SQYGGASAPVIRGTGRIKVLNHHGTGMDAFSPDHAIWDTALSOQVEILRGVPTL 145
Db 89 NQFGGSGAPVVRGQGVRLKTLQNGSDVIDMSQLSPDHAIGVDTLLAEQVEIVRGASTL 148
QY 146 LVSSGNVAGLVADVADGKIPEKMPENGVSGLGLRLSSGNLEKLTSGGINIGLGNFVLHT 205
Db 149 LYANASPAVINVDKRIPTQPKGYEVDFNTRYNTNSHEKLVTAALTFGLGHIALRV 208
QY 206 EGLYRSGDYAVPRY---RNLRKLPDPSRPFANGQHRVAVLGRKRFYRTYSDRRDQVGL 262
Db 209 BELLRGSNNYHVPFALDKTLNVPDTQNKTKSGNYGVAFTGERGYGFAYNLRREKYGL 268
QY 263 PAHSHEYDDCHADIIWQKSLINKRLVQLYPHLLTTEEDVDYDNPGLSCGF-HDDDDAHAHA 321
Db 269 FGHNHKLDSCAAH-IWGNVNRNDYLYGLYPHLM--HDTDLVNTPHCGSNHMDGKSHNN 325
QY 322 HN-----GKPTIDLENKEYELRAEWKQFPFGFEALRVHLNENDYHDEKAGDAVEN 372
Db 326 HPYGHDDHDSIAGPLIDSYAKRYDRAEVKQPMKAIEKILSYSETRYKHKDEKGNIAVN 385
QY 373 FPNQTONARIELRHQPTGLRKGSGVQVYLGQSSA----LSATSEAVKQP----- 419

```

```

Db 386 LFKNGYNLRVIEIRHTPIAGLSGVIGAQYQOTQSSANIPRIAPCSNNASDPCHKKKORDP 445
QY 420 -----MLLDNKVQHYSPFGVQANWDNFTLEGGYRVKQKASIEYDKALI----- 464
Db 446 SKITGDKRKSALLIENTSQSQMSFAIEQLRWQDFLFEIGVTEKQKRIEYDRAWLEFKVK 505
QY 465 -----DRENYI-----KOPLPDLAGHROTARSPALSGNMYFTPOHKLSLTASHQE 509
Db 506 KLEGCDFNSFFYSPSGCRQGSYPAPDFASVHTRATSYSGAISWNMTPTDYLSTLYSHNE 565
QY 510 RLPSTQEILYAGKGVAVNTTFVGNKHLNKNERSNNIELALGYEGEDRWQYNLALYRRCGY 569
Db 566 RHPTPMELYHGHKHLATVSEFHGNRLKKEYSDNWEVGLAYLGDLSYKXNVVYNDPKNR 625
QY 570 IYAQTLNDGRGPKSIEDDSEMKLVRYNOSGADFYGAEIEYFKPTPRYRIGVSGDYVRGR 629
Db 626 IFNQTLN-----KSNLSLRYNOSKAKYIYQVEGRIDYALTPELHMGILFGDYVRGK 676
QY 630 LKNLPSLFGREDAYGNREFIAQDDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVFAQN 689
Db 677 LYDLPTVYRVHDVANSLEPVTQPDQAPRVPEMRLGFRVNMMENTESLTSLEYTYVYQK 736
QY 690 KLARYETTPGHMLNLCANVRNTRYGEMWVYKADNLNQSVVAHSSFLSDTPQMGRS 749
Db 737 KVAPLENQTAAYSLNIGVYSRQIAGVNYQLFVQANNVLRKRYSHTSFLFPFPMGRN 796
QY 750 FTGGVNVKPF 758
Db 797 VTGLGNIHF 805

RESULT 6
Q9CMR7 PRELIMINARY; PRT; 925 AA.
ID Q9CMR7;
AC Q9CMR7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein PM0745.
GN PM0745.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006112; AAK02829.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep Rec; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 925 AA; 105081 MW; F46A9F52B59145E3 CRC64;

Query Match 32.4%; Score 1307; DB 16; Length 925;
Best Local Similarity 33.7%; Pred. No. 2.1e-85;
Matches 292; Conservative 149; Mismatches 278; Indels 148; Gaps 14;

QY 29 QSVGLETVTVGKSRPRATSGLLHTSTASDKTIISGDTLRQKAVNLGDALDGVPGIHA 88
Db 70 QTEMLDEIVVSGAQPHLAGSAIEHYQAIISNQVKKERLQKQSATLGNALAGELGVHSNFF 129
QY 89 GGASAPVIRGTGRIKVLNHHGTGMDAFSPDHAIWDTALSOQVEILRGVPTL 148
Db 130 GGASAPIIRGQGVRIKILQNGLDVYDMSAISPDHAAVADSLAEQVELRGASTL 189
QY 149 SGNVAGLVADVADGKIPEKMPENGVSGLGLRLSSGNLEKLTSGGINIGLGNFVLHTGL 208
Db 190 SASAAGVNVVDKRIETAVFEKGYEGEITFRDFTASQESTGTAGITFRHLPHLALALESL 249

```

```

QY 209 VRKSGDYAVPRYN---LKRPLDPSRRPANGHRAVLGWRK--RPRRTYSRRDYOVL 262
Db 250 KRYSTHYRVPAPKSGETIRYLPDSHNRSGVG--TIGVSWIKDQSYGLVGSYSERRDRYGL 307
QY 263 FAHSHEYDDCHADIIWQ--KSLINKRYLQLYPHLLTTEEDVDYDNPGLSCGFFHDDDDAHAAH 320
Db 308 PGNHKKYDRCKSHVDEAARPELGGKYLTPYHLADDDTDIVAHLDGCGIGGIDNDFSHSH 367
QY 321 AH-----NGKPWIDLNRKRYELRAEWKOPFGFEALRVHLNRNDY-HHDEKAGDA- 369
Db 368 DHFPFGHEHDSHGGFWRLHSRFDLRGQWDSPTAWLDKVKGSFSYADYIHYEYHSGQAG 427
QY 370 -----VENFFNNQONARIELRHOPIGELKSGWGVYLGQKS 406
Db 428 TKDFRDSFIERERKAEKVRKAAGIYKNSGYNGKLEFYHTPIVGLSGVFGVQVSEYKT 487
QY 407 SALSATSEAVK-OPMLLDNKVQHYGFFGVEQANWDFNLTLEGGVRVEKQKASIRYDKALID 465
Db 488 AILAPLGSIGIKHQHLLVPNTQAGSFAVENYVDDDFEFGVARDVKQRIPIKYDQHVLN 547
QY 466 RENYYKQPLDLAGHRTARSFALSGNWFYTPQHKLSLTASHOERLPTSTQELYAHGKHVA 525
Db 548 AHKKGDDHPPDLSTHKEKAVSLGSDVWDLFHPNYRIGLTLRNERLPTPMELYHYHQHLLA 607
QY 526 TINTFEVGNKHLNKRSSNNIELALGYEGDRQWQYNLALYRNRFGNYIYAOTLNDGRGPKSTE 585
Db 608 TNSFEHGNKLRKESNNIELGFAHTDWDYKLSLYQNKFRNYLYNEDL--ARYGNAP- 664
QY 586 DSEMKLVYRNSGADFYGABGEIYFKPPTRYRIGVSGDYVRGRLKNLPSLFGREDAYG- 644
Db 665 -----LRRYTQARAKFHGIEAELNFRPTPDYQVTLFGDYVRGRLFDLPEQYQRFYQY 718
QY 645 -----NRPF-----IAQDDQNAVRPAARLGFHLKASITDRIDANLDYRVF 686
Db 719 IAYDEGLQANWEKQPIYRIGIKRNERDAPRVPARLGMRLSGNVTEHLSFFADYTYVF 778
QY 687 AQNKLA-----RY----- 694
Db 779 SQOKTASSLSIKPRALEASDFIDDTGENTLLKGIDRKYNRTQPTADLSASDKADPE 838
QY 695 -----FRTPOHMLNIGANYRNRTRYGEMWYVYKADNLLNQ 731
Db 839 SALEAIRLEENKLPKATIEKIQEDPSKGHLLNIGVNYQRIQIHLDSYSGVSVANNVQ 898
QY 732 SYVASSLSLDPQMGSRSTGVNWK 758
Db 899 RVVHTSYLPPVQMGRRNYVNLNFGVKF 925

RESULT 7
Q8P7F3
AC Q8P7F3 PRELIMINARY; PRT; 688 AA.
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Outer membrane hemin receptor.
GN PHUR OR KCC2658.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Camarotte G., Cannamaro F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

```

```

RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C. de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT *Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.*;
RL Nature 417:459-463(2002).
EMBL: AE012378; AM41930.1; --
InterPro: IPR000531; TonB boxC.
DR Pfam: PF00593; TonB dep Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 688 AA; 74262 MW; 2E47F6BB47F6429B CRC64;

Query Match 18.5%; Score 745; DB 16; Length 688;
Best Local Similarity 27.5%; Pred. No. 5e-45;
Matches 219; Conservative 123; Mismatches 301; Indels 152; Gaps 21;

QY 1 MAQTTLKPTVLISILLINTPLLAQA-----HETEQSVG-----LETVT 37
Db 9 LRERSTLTALASVLI---PALAMADDTVPADPQTPPSSDTRDATHATPPSGRHIKDLKV- 64
QY 38 VVGKSRPRATSGLLHTSTASDKIISGDTLRQ-KAVNLGDALDGVPGIHAHQYGGGASAPV 96
Db 65 VVTASPLRDAAGELSRPV---EVLAGERLDEVRASSIGETVASLPGVQSSNFGPGVGRPI 121
QY 97 IRGQTRRIKVLNHHGEICDMADFDPDHAIMVDVTALSOQVEILRGVPTVLLYSNGNVAGLV 156
Db 122 IRLGLDGPVAVLRDLGSLTQDVSTVSQDHSIPAELFANQIEVLKGPSTLLYSNGAIGGVV 181
QY 157 DVADGKIPEKKPENGVSSELGLRLSSGNLEKLTSGGINIGLGKFNVLHTEGLYRKSQY 215
Db 182 NVVDGRIAE-TPVDGFSGEAEVRFDDGKDGNTDMFVDVAGNGSGSLSIHADGVYRNQDY 240
QY 216 AVPRYRNLRKLPDPSRRPANGHRAVLGWRKFPYRTYSRRDQYGLPAHSHEYDDCHAD 275
Db 241 DTPQGRQANSFLDS---KVGSIASLAGDMGFWGLSASRFRDNYGNPGE----- 286
QY 276 IHWKSLINKRYLQLYPHLLTTEEDVDYDNPGLSCGFFHDDDDAHAAHNGKPMIDLNRKY 335
Db 287 -----PGDLSIGERGS-----LKLQDDRY 306
QY 336 ELRAEWKOPFGFEALRVHLNRNDYHDEKAGDAVENFFNNQONARIELRHQPIGRKLG 395
Db 307 DLKGLGLTPWGDGSLRYSFGHTDYAHTFEAGAEVGTFTKRANEGRVEASTFGGQWOT 366
QY 396 SWGVQYLQOKKSALSATSEAV-KQPMLLDNKQHYSFSGVEQANWDFNLTLEGGVRVEKOK 454
Db 367 AFGLO-----GSDTTFQAVGEESFVPKTDTRSLGLFGVARNSDRVTAEVGARVDVKV 419
QY 455 ASIRYDKALIDRENYKQPLDLAGHRQ-TARSFALSGNWFYTPQHKLSLTASHOERLPS 513
Db 420 YS-----TDIGVDRDFTPTFSASAGFRNEQWRLTANLDHARAPA 461
QY 514 TQELYAHGKHVATNTFEVGNKHLNKRSSNNIELALGYEGDRQWQYNLALYRNRFGNYV- 571
Db 462 EEELFANGPHATLAFELGIDANILKTEKANQALGNLFQNDWDVDAKIAAYNRYNDYFVIV 521
QY 572 ---AQTLDNRGPKSIEDDSEMKLVYRNSGADFYGABGEIYF-----KPTPRYRIGVSGD 624
Db 522 DTGGQWFNE-----EDNDFLPIRQWTAQDAIFHGPEGEATFHLANNDDTGAWDLRVFGD 574
QY 625 YVRGELKNLPSLPGREDAYGNRPFIQDDQNAVRPAARLGFHLKASITDRIDANLDYR 684
Db 575 TVQARLK-----DGGNLPRIAPGRVGAEMRW-N-ADAWRASLGAI 613
QY 685 VFAQNKLARVETRTPGHHMLNIGANYRNRTRYGEMWYVYKADNLLNQSIVYHSSFLSDTP 744
Db 614 TMKQKVAVNETATDGYTMDVAHLAYHMDRGANAEVFLDGNLNTNQDARVHTSFLKDDV 673

```

```

QY 745 QM-GSFTGCVNKF 758
DB 674 MLPGRNASFGVRLFF 688

RESULT 8
QSP16 PRELIMINARY; PRT; 687 AA.
ID QSP16 PRELIMINARY; PRT; 687 AA.
AC QSP16 PRELIMINARY; PRT; 687 AA.
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Outer membrane hemin receptor.
GN PHUR OR XAC2829.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Perzo J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AB011924; AAC37674.1; -.
DR InterPro; IPR00531; TonB boxC.
DR Pfam; PF00593; TonB dep.Rec; 1.
KW Receptor; Complete Proteome.
SQ SEQUENCE 687 AA; 74280 MW; A0B719C4ED4F6D1F CRC64;

Query Match 17.8%; Score 718.5; DB 16; Length 687;
Best Local Similarity 27.1%; Pred. No. 4.1e-43;
Matches 213; Conservative 122; Mismatches 301; Indels 149; Gaps 21;

QY 11 LGLILLNTPLLAQA-----HETEQSVG-----LETVTTVGKSRPRAT 47
DB 15 LALASLLTPALAWEDAPVAADGQTPSPSDTHDASPSSGRHDKDVK-VTASPLDA 73
QY 48 SGLLHTSTASDKIISGDTLRQ-KAVNLGDALDVPFIHASQYGGASAPVIRGTGRK 106
DB 74 AGELSRPV---EVLAGERLDEVRSSLSGETVASLPGVQSSNPGVGRPIIRGLDGRVA 130
QY 107 VLNHGEGTDMADFPDFAIMVDYDTSQVETLIRGVITLYSSGNVAGLVDAVKIPEK 166
DB 131 VLRDGLSTQDVTQVSDHSPALPEPLAQIIEVLRKGFSTLLYSGAIGGVVNVVGRTAE- 189
QY 167 MPENGVSGLBLRLSSGNLEKLTSGGINIGLKKFVLHTEGLYKRSQGYAVPRYRNLKR 225
DB 190 TPVDGFSGRAEVRFDGDKDGTDMFRVDAGNSGLSIHADGVYRNQNDYTPKGRQLNS 249
QY 226 LPDSPRRFANGHRAVLGWRKPEYRTYSDRRDQGLPAHSHEYDDCHADIIWQKSLINK 285
DB 250 WVDSD---KVGSGISLSDGWDGVLGSASRFRDNYGNFGE----- 285
QY 286 RYLQLYPHLLTEEDVDYDNPGLSCGFPHDDDDAHAAHNGKWPIDLNRKRYELRAEWKOPF 345

```

RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira M.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubaki M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AE003890; AAF83194.1; --
DR InterPro: IPR000566; Lipoclin_cyFABP.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Rec; 1.
DR PROSITE: PS00213; LIPOCALIN; 1.
KW Complete proteome.
SQ SEQUENCE 681 AA; 74084 MW; E7FFDF36F1CEE5ED CRC64;

Query Match 17.2%; Score 694.5; DB 16; Length 681;
Best Local Similarity 27.4%; Pred. No. 2.2e-41;
Matches 220; Conservative 115; Mismatches 294; Indels 173; Gaps 22;

QY 1 MAOTTLKPIVLSILLINPTLLAQA-----HETQSVGLFTVTVVG----- 40
DB 9 LPRTVFLAVSLL--TPALAMADVGPPTDSGRHFKTTEVTRHLKDLDAVVTAIPLRD 66
QY 41 ----KSPRATSGLLHSTASDKLIISGDTLRQ-KAVNLGALDGVGIIHASQVGGASAP 95
DB 67 SYDLSRPVA-----LLAGERLDEVRASSLGSETAVLPVQVSNFGPGVGRP 113
QY 96 VIRGQTGRRIKVNLHGHETGMDSPDHAIMVDLTALSOQVEILRGPTVLLYSGNVAGL 155
DB 114 IIRGLDGPRIAVLNGLSSQDVSTVSDHSPAVEPFLANQIEVLKGPSTLLYSGAIGGV 173
QY 156 VDVAADGIKPKMPNGVSGELGLRLSSGNLEKLTSGGINIGLGNFVLTGELYKSGD 214
DB 174 VNIVDGRIAE-APYGVGFNGRAEMRLDGGKHGNTNMFRI DAGNGSALSVAHADGYRNEK 232
QY 215 YAVRYRNLKLPDSRRFANGQRAVLGWRKRYRRTYSDRDQYGLPAHSHEYDDCHA 274
DB 233 YDTFGQVNSFIDT----KSGSWGASFSGDFGFGLSVARPHDSYG----- 275
QY 275 DIIWQKSLINKRYLYPHLLITTEEDVDYDNPGLSCGPHDDDDAHAAHNGKPIDLRNK 334
DB 276 -----NFG-----EPGDVPAGDRGSLRLHQDR 298
QY 335 YELRAEWKQPPGFEALRVHLNRNDYHDEKAGDAVENFFNNOFONARIELRHQPTGRK 394
DB 299 YDLKAGLTDPGESALRPFSLCHTQYDIEPEGNEVGTTFKRASEGRVASTAFGGWR 358
QY 395 GSWGV-----QYLGQKSSALSATSEAVKQPMLLDNKVOHYSFPQVEQANWNTFLEGG 447
DB 359 TAFGVQGGDSTFQALGEEFVPEKTNKSI-----GVFGLAHNTFGLFQAEFG 405
QY 448 VRVEKQKASIRYDKALIDRENVYKQPLDGLGAHQRTARSFALSGNWYFTPHKLSLTASH 507
DB 406 ARGDQ-----VKYTDNGVTRNYH-----PGSLAFSGDLALSQKWLRLTNVDH 448
QY 508 QERLPSTQELYAHGKHVATNTFEVGNKHLNKRNNIELALGYEGDRQWYNLALYRNRFG 567
DB 449 AERAPVEEELFAKGPHIATLAYEVRGADLKEKANAQELGLVFRNDNSDAKVSITYSYG 508
QY 568 NYIYAQTILNDGRGPKSIDDESEMILV--RYNSGADFYGAGEIYF-----KPTFRYIGV 621
DB 509 NFIV---LVD-TGTFWDDQORDLFRVQWSEQANAI FHIIEGEATFHLAKNTSGSDNLRV 564
QY 622 SGDYVRGLKXNLPSPGREDAYGNRPFTIAQDCQONAPRPAARLGFHLKASITDRIDA--- 678
DB 565 FGDTVSGRLKN-----GGNLPRIVPARYCAELRWE-----DAGWR 599
QY 679 -NLDYRYPQAOKLARYTRTPGHMLMLGANYRNTRYGWNVYKADNLLNOSVYH 737
DB 600 TSLSAKRYEKQNKVAVNETPTAGYTWDAHLAYHIDVDSIAWEVFFDGNLNTNRDARVHT 659
QY 738 SFELSDTPQM-GRSFTGGVNVKF 758

DB 660 SFLKDDVMLAGRNYTAGLRMEFF 681

RESULT 10
Q915F7 PRELIMINARY; PRT; 687 AA.
ID Q915F7;
AC Q915F7;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein PA0781.
GS PA0781.
ON Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A. / PA01;
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garbay R.L., Coulter L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Goltz S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saiter M.H., Hancock R.B.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004513; AAG04170.1; --
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Rec; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 687 AA; 74205 MW; FBEC6ACD94B8FCD CRC64;

Query Match 16.8%; Score 679; DB 16; Length 687;
Best Local Similarity 27.4%; Pred. No. 2.9e-40;
Matches 212; Conservative 121; Mismatches 319; Indels 122; Gaps 25;

QY 6 LKPIVLSILLINPTLL-QAHETEQSVGLFTVTVVGVKSRPRATSGLLTSTASDKLIISGD 64
DB 15 LTPLALACIIVSGETIGADGRPSLPQSVITANPLGNESPATPS-----SVLEGD 64
QY 65 --TLQKAVNLGDALDGVPGIHASOYGGASAPVIRGOTGRRIKVNLHGHETGMDAFSP 122
DB 65 EHTLQKQ-SLEGTNLGLPGVSTTFPGASRPVIRGMDGRIRLNRNGVGLDASSLSY 123
QY 123 DHAIMVDTALSOQVEILRGPTVLLYSSGNVAGLVADGKIPEKMPNGVSGELGLRLSS 182
DB 124 DHAVPEDPNSVERLEVRGPAALYGGNAIGGVNSFDNRIPSE-PVDGIHSGELRYGG 182
QY 183 GNLEKLTSGGINIGLGNFVLTGELYKSGDYAVRYRNLKLPDSRRP--ANGQHRA 240
DB 183 ADTTRSRSGALEAGDG-NFALHVDAAAREFDNRVIRPGVAHSR-----QRIDGTGTRHV 237
QY 241 VLGWRRKRYRRTYSDRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLYPHLLITTEEDV 300
DB 238 -----QNSDGRQDGGAVGSGYHWEHGYAGLSY-----SGYDS 269
QY 301 DYDNPGLSCGFHDDDDAHAAHNGKPIDLRNKRYELRAEWKQPPGFEALRVHLNRNDY 360
DB 270 NYGSPA-----EDVR-----LKMQDRYAFASEIRDLLEGPTSLKDAAYTKY 313
QY 361 HHDEKAGDAVENFFNNOFONARIELRHQPIGRKLSGWVQYLGQKSSALSATSEAVKQPM 420
DB 314 EHKELIDGEGTTFKNEGYEGRIEHRPLGLPLNGVGAQFANSRFSALG-----EAF 367
QY 421 LLDNKVQHYSPFGVQANW---DNFTLEGVREVEKQKASIRYDKALIDRENVYKQPLD 477
DB 368 VEHTTDSALFALEE--WKLSDRLDLSFGARLE---HTRVDPDAKNGERFAEND--- 417
QY 478 GAHQRTARSFALSGNWYFTPHKLSITASHOSELSTQELYAHGKHVATNTFEVGNKHLN 537

```

Db 418 GSQFTTGSSTGAVYKLTPIWSLAATLSYTERAPTFYELANGPHAATGYEVGDAD 477
Qy 538 KERNNIELALGYEDRWQYNLALYRNRFNGYIYAQTLDNGRGPK-----SIBDDSEMK 591
Db 478 KEKAVSTDLALRFDFGVHKGSGVGFYSRFSNYI--GLIASGRHNEEGEVVAAGDDEALP 535
Qy 592 LVRNQSGADYFGAGE-----LYPKPTPRYRIGYSGDYVGRGLKNLPSLPREDAYGNRP 647
Db 536 EYLYSGVWRADYFGEAQDRHLLSPYGNFDELSGDYTRAKND-----580
Qy 648 FIAQDDQAPRPAARGLFHLKASITDRIDANLDYRVFAQNKLARVETRTPGHMLNLG 707
Db 581 ----TGEPLRIAPLRUNLALIWEL--QQQARVDVEHAASQHRVPEELSDGY--TTLG 633
Qy 708 ANYERNTRYGWNW--YKADNLLNQSVYAHSSFLSD--TPQMGSRFTGGVNVKF 758
Db 634 ASLGYNFDLGSRLAFYKGTNLNTQTVRYASSILDRVPAAGRGIEAGVKVAF 687

RESULT 11
Q9ABL3
ID Q9ABL3 PRELIMINARY; PRT; 687 AA.
AC Q9ABL3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE TonB-dependent receptor.
GN CC0214.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Winn M.B., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005695; AA22201.1; -.
DR TIGR; CC0214; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 687 AA; 72786 MW; 29415A726A3BF753 CRC64;

Query Match
Best Local Similarity 15.8%; Score 636; DB 16; Length 687;
Matches 206; Conservative 107; Mismatches 237; Indels 168; Gaps 23;

Qy 22 AQAHETEQSGLTETVVGKSRPRATSGLLHTSTASDKIISGDTLRQKA-VNLGDALDGV 80
Db 37 ASAQDAEVSKVIITAPYGVY-----ADALTASVAVLDRTDLDLAAPKGLDALAG 88
Qy 81 PGHSAQYGGGASAPVIRGTGTRIKVNLNHGFTGDMADFPDHAINMVDLTALSQQVEILR 140
Db 89 PGVRSSTTFGAGSRPVVRGLAGPRVQVLTNGVGQIDASALSPDHQVATDPGEAERIEVL 148
Qy 141 GPVTLIYSSGNVAGLVADVADGKPEKMPENGVSGELGLRSSGNLEKLTSGGINIGLGN 200
Db 149 GPAALAYGSAIGVNNIIDRISTQOPIDGMSGRLLASRGTDGSDSALSAGVHATVGP- 207
Qy 201 FVLHTEGLYKSGDYAVPRYNLKLPLDSPRFA-----NGQHR-----AV 241
Db 208 MVLTLDAKKEESDKYIPVY-----PESARQALAEGETAEGAGRLSENSAVDLETFGAG 261
Qy 242 LGW--RKRFRYRTYSRRDQYGLPAHSHVEDDCHADIWQKSLINKRYLYQLPHLLTEED 299

```

```

Db 262 LSYVDKRGFVGMSIKRTDSTGYVGPCHAHE-----290
Qy 300 VDYDNPGLSCGFHDDDDAHAAHANGKPMWIDLRNKRXYELRAEWKQFPFGFEALRVHLNRND 359
Db 291 -----HEGEABAGHEESAVTIGLUKQTRIDLRGEYDADLGGFAKVRFGGGHAD 338
Qy 360 YHHDKAGDAVENFNQNTQNAIRIELRHQPTIGRLKSGWGVQVYLGKXSALSATSEAVKQP 419
Db 339 YTHTEFGDAGVTRFTSDGYEGRLVQTERGGWKAGVGVQALARNPDAIGEAIVPK--396
Qy 420 MLLNKVQHSYFFGVQGANWDFLEGGVRVEKOKASIRYDKALIDRENYKQPLPDLGA 479
Db 397 ----TKITEFGAFTQORLELDGYEGGLRITDRELDSLKGKA--DFTN-----439
Qy 480 HRQTARGFALSGN-----WYFTPOHKLSTASHQRLPSTQELYAHGKHVATNTFEVGNKH 535
Db 440 -----LSGSVGAFAWRPTTESFVGLALSRSRAPTESELFAGPHAATRGFEIGDAD 490
Qy 536 LNKERSNIELALGYEDRWQYNLALYRNRFNGYIYAQTLDNGRGPKSIEDDSEMKLVRY 595
Db 491 LKEETATSLATLHYGGERVSGDLHLVIARYDGFIDLRPTGD-----EEDG-LAVRY 542
Qy 596 NQSGADYFGAGEIYFPTPRYRIGVSG-----DYVRGLKNLPSLPREDAYGN 645
Db 543 VQTDABEFGFEAEI-----AYRVWTDGQSRVNLHAGADFVRG-----579
Qy 646 RPFIAQDDQAP-RVPAARGLFHLKASITDRI--DANLDYRVFAQNKLARVETRTPGHHM 703
Db 580 -----SSDLGPPARIPYVS--LKASYQAPMWSGDVEVRTTGGQERVAQNELPTDGYTV 632
Qy 704 LNLGANYR--RNTRYGEMWNYKADNLLNQSVYAHSSFLSD--TPQMGSRFTGGVNVKF 758
Db 633 LNAATLAWKPAGNARV--RLFLDGRNLTNEEVREHVSFLKDIAPSPGQVRAGIALRF 687

RESULT 12
Q8F1I6
ID Q8F1I6 PRELIMINARY; PRT; 777 AA.
AC Q8F1I6;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hemin receptor.
GN IA3149.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE011477; AA050347.1; -.
KW Receptor; Complete proteome.
SQ SEQUENCE 777 AA; 86615 MW; 03F985150B36589A CRC64;

Query Match
Best Local Similarity 9.8%; Score 395.5; DB 16; Length 777;
Matches 173; Conservative 121; Mismatches 305; Indels 169; Gaps 29;

Qy 28 EQSGVLETVTVGKSRPRATSGLLHTSTASDKIISGDTLRQKA-VNLGDALDGVPGIHAS 86
Db 105 ESSLDQSAINTAKS-----TNSDFLTAPQPIITVLSGRQLDRQGETAMSAINTTGVSNL 160
Qy 87 QYGGGASAPVIRGTGTRIKVNLNHGFTGDMADFPDHAINMVDLTALSQQVEILRGPVTL 146
Db 161 TTSGSGTKPIIRGLTQORVLMVT--DGIRQEBEQFGDDHTVELDSFNKIEIRPGSL 219
Qy 147 YSSGNVAGLVADVADGKPEKMPENGVSGELGLRSSGNLEKLTSGGINIGLGNFVLHTE 206
Db 220 YGSDALGGVNVIRDKAP-----LSGE-----GIPKAGIFNS 252

```


Query Match	6.6%; Score 264.5; DB 16; Length 719;	
Best Local Similarity	22.5%; Pred. No. 2.7e-10;	
Matches 169; Conservative 118; Mismatches 328; Indels 137; Gaps 38;		
QY	11	LSLLINTPLLAQAHETEQSGVLETVVVGKSPRATSGLLHTSTASDKIISGDTLRQKA 70
Db	19	LSLCISGIVSTATATSEYTKISNEETLVVTTNR---SASNLWESPATIQVDDQTQNST 75
QY	71	-VNLGDALDGPVGIHSAQYXG-GGASAPVIRGQTRRIKVLNHHGE-TGDMADPSPDHAIM 127
Db	76	NASIAADNLQDIPGVEITDMSLAGRKQIRIRGEASSRVLLILDQEVTYQRAAGNYGVGLL 135
QY	128	VDTALSQQVEILLRGPVTLLYSSGNVAGLVADGKIPKMPENGVSGELGRLSSGNLEK 187
Db	136	IDESALERVEVVGKPSVLYGSAIGVINFTIKGGDKL-ASGVVKAIVNSATAGWEES 194
QY	188	LTSGGINIGLKNFVLHTEGLYKSGDYAVPRYRNKRLPDSRRFPANGQHRVILGWKR 247
Db	195	IAVQG-SIG---GFDYRINGSYSDQNRDTPD---GRLPNT--NYRNSQGVWLGVNSG 244
QY	248	FVRTYSRRDQVGLPAHSHYDDCHADIWQKSLINKRYLQLYPHLLTEE-----DVDY 302
Db	245	NHR--FGLSLDRYRLATQTY-YEDPDGSY-----EAFSVKIPKLEREKVGVFYDTPV 293
QY	303	DNPGSCGFHDDDAHAHANGKPMWIDLRNKRVELRAEWKQPPGFEALRVHLNRNDYHH 362
Db	294	DGYLK-----KHFDAVEQTIQRFAN---EVKTQPVPSFMIQALTVH-NKTD--TH 341
QY	363	DEKAGDAV--ENFPNNQTONARIELRHQPIGRILKSGWGVQY-----LGQKSSALSATSEAV 416
Db	342	DKQYTOAVTLQSHFSLPANNELVT-----GAQYKQDRVSRSGMGTSSKSLT 388
QY	417	KQMLLDNKVQHYSPFGVQQA-----NW-----DNFTLEGVVRVEKQKASIRYDKALID 465
Db	389	G---FINKETRTSYYESQSTVSLFAQNDWQFADHTWTMGVR--QYWLSSKLTRG--D 441
QY	466	RENYKQPLDLAGHROTA--RSFALSGNWYFTPOHKLSLTASHQER--LPSTQELVYAHG 521
Db	442	GVSYTAGIISDLSLARESADHEMTVTSLSRYSGFDNLELRAAFAGYVFPFLUSQLFMQT 501
QY	522	KHVATNTFVGKHLNKRNSNIIEALGYEGDRWQYNLALYRNRFNGYIYAQTLNDGRGP 581
Db	502	SAGGSVTV--GNPDLKAHSNNPELGARYNGNQWLIDSAVYSEAKDYI-ASLICDGSIV 558
QY	582	KSTEDSEMKLVRY--NQSGADFYGAGEIYFKPTPRYRIGVSGDYVRGRL----- 630
Db	559	CNGNTNRSRSGYYYNDIDRAKTWGLE-----ISAENGWVFPSPYISGNLIRROYETST 612
QY	631	-----KNLPSPGR-----EDAYGNRPFFIAQDDQNA--PRVPA-ARLGPH 667
Db	613	LKTTNTGEPAINGRIGLKHLYVMQANIISDVFIRAASSAKDDSNGTETNVPGWATLNA 672
QY	668	LKASLTDRIDANLDYRV-FAQNKLA--RYET 696
Db	673	VNTEF-----GNBDQYRIINLNLNLTDKRYRT 699
RESULT 18		
ID	Q9CN63	PRELIMINARY; PRT; 742 AA.
AC	Q9CN63;	
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	HemR.	
GN	HMR OR PM0576.	
OS	Pasteurella multocida.	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;	
OC	Pasteurellaceae; Pasteurella.	
OX	NCBI_TaxID=747;	
RN	[1]	
RP	SEQUENCE FROM N.A.	

RC	STRAIN=Pm70;	
RX	MEDLINE=21145866; PubMed=11248100;	
RA	May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;	
RT	"Complete genomic sequence of Pasteurella multocida Pm70.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).	
DR	EMBL; AE006093; AAK02660.1; -	
DR	InterPro; IPR000531; TonB_boxC.	
DR	Pfam; PF00593; TonB dep Rec; 1.	
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.	
KW	Complete proteome.	
SQ	SEQUENCE 742 AA; 84910 MW; 9B2814BDAAD229D1 CRC64;	
Query Match		6.3%; Score 253; DB 16; Length 742;
Best Local Similarity		20.8%; Pred. No. 2e-09;
Matches 165; Conservative 117; Mismatches 291; Indels 222; Gaps 39;		
QY	67	ROKAVNLGDGPVGIHSAQYGGASAPVIRGQTRRIKVLNHHGETGMDAFS----- 121
Db	61	QQQADNAADLINILFGVNA-----GGFRP--GGQT-----LNING-MGDAEDRVQLDG 107
QY	122	-----PDHAIMVDTALSQQVEILRGPVTLLYSSGNVAGLVADGKIPKMPEN-GV 172
Db	108	ATKSPKYOQSGSIFIEPELLRKVTVDKGNVSPQYGNNGFAGTVKFTKDATDFLKNQKI 167
QY	173	SGELGRLSSGNLEKLTSGGINI-GLGKNFVLHTEGLYKSGDYAVPRYRNKRLPDSPR 231
Db	168	GGLFKYGNNSNNQKTYSTALVLQNEQKNIIDLLFGSVRNASNYTRP---DKSKILFSKN 224
QY	232	RFANG-----QHRAVL-----GWRKFRYRTYSRRDQVGLPAHSHYDDCH 273
Db	225	NQKSLGIKVNWQITPEHLLTLFSVGIHKGW-----EPWAAKEDVMSRPTET-EIKRYG 277
QY	274	ADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHANGKPMWIDLRNK 333
Db	278	IDVAKRKLIVR-----DQKDESYSL-----KYRYPENNNKWINLSVQ 315
QY	334	RYELRAEWKQPPGFEALRVHLNRNDYHDEKA-----GDAVENFNNQT 378
Db	316	LSYSKTE-----QNDTRHEKVTSSFLGTIGNKSWITYSDLTIDISTST 359
QY	379	QNARIELRHQPIGRILKGS--WGVQYLGQKSSALSATSEAVK-----QPMLLDNKVQ 427
Db	360	LN-----IGRAHELLFGLQWLKNKRNITLVYHKEGVKKADYNGYFQPYMPSGRQ 410
QY	428	H-YSPFGVQANWDNPTLEGGVRVEKQKASIRYD-----KALIDRENYKQPLDPLGA 479
Db	411	YTHAFYLDQIKWQNFLEFTGG-----IRYDHINNIGQKNLAPRYN-----DISA 454
QY	480	HQOTARSFALSGNWY---FTPOHKLSLTASHQE--RLPSTQELYA--HGKHVAINTPE 530
Db	455	GHDYSQKNTNGWSYILGLKAYDVNHYLSLFTNFSKTWRAPVIDEQETQYSQASVSATSLN 514
QY	531	VGNKHLNKRNSNIIEALGY--EGDRWQYNLALYRNRFNGYIYAQTLNDGRGPKSIEDD 587
Db	515	LEKEMINQTRVGGI-ITLNLHLPQENDAFQFRITTYFYNRGKNEIFKT-----RGVNCVENA 568
QY	588	SEM-----KLVR--YNQSGADFYGAGEIYFKPTPRYRIGVSGDYVRGRLKNLPDLGR 639
Db	569	LVDVNSVCPKIIISNRYNRLPGYVIQGALEAYYQSSYLFQ-GLTYSYVKGKRDTSFRNPS 627
QY	640	EDAYGNRPFFIAQDDQNAQPRVPAARLGFHLKASLTDRIDANLDYRVVFAQNKLYARTRP 699
Db	628	KTS-----TWIA---ETPPRKATATLGFNIPEY-----YFTAGWRAEFVRKQDRSP 670
QY	700	-----GHHMLNLGANYRNTRYGEW-----NWYVKADNLLNQSVYAHSSFLSDTP 744
Db	671	LSNDSKASYWALPSSKGYSLHSVFSWSPTKIKGNFKVTVDNLFNRPYY-----P 721
QY	745	QNGRSFTG-GVNVKF 758
Db	722	YLGELASGTGRNVKF 736

RESULT 19
 Q9HYX3 PRELIMINARY; PRT; 721 AA.
 AC Q9HYX3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Probable TonB-dependent receptor.
 GN PA3268.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Gotlib L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Lazbig K., Lim R.M.,
 RA Smith K.L., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sauer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 RT opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 RL EMBL; AE004749; AAC06656.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR InterPro; IPR006195; tRNA_ligase_II.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00862; AA_TRNA_LIGASE_II; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 DR Receptor; Complete proteome.
 SQ SEQUENCE 721 AA; 79371 MW; 70B6716698A9E05D CRC64;

 Query Match 6.1%; Score 246; DB 16; Length 721;
 Best Local Similarity 21.2%; Pred. No. 6e-09;
 Matches 184; Conservative 102; Mismatches 277; Indels 304; Gaps 42;

 Qy 4 TTLKPIVLSILLINTPLLAQAHETEVSGLTETVTVGKSRPRATSGLLHTSTASDKII-S 62
 Db 5 SSLSRLAVALAIGVSMQAQAEEAEKELG--TIVTVGDLGEADQAVVQNHGARSVVR 62

 Qy 63 GDTLRQKAVNLGDALDVGPHGASQYGGGASAPV-----TGTGTRIKVNLNHHGETGDM 117
 Db 63 REMLESQAQNRVRLKVPQVQVQDNNNGTGGSDISLNVGVRGLTSR----- 108

 Qy 118 ADFSPDHAIWVD-----TALS-----QQVELRGFTVLLYSSGNVAGLVD 157
 Db 109 --LSPRSTVWLDGVPAAVAPYQPOLSMPLSIGNLESIDVVRGAGSVRYPQNVGVIN 166

 Qy 158 VADGKIPEKMPENVSSELGLRL---SSGNLEKLT---SGGINIGLGRNFVHLHTEGLYR 210
 Db 167 FVTRAIPEKF-----SGEIGTTIEHAGGKWKLNQAFGLGTADNGLG--VALLYSV-- 217

 Qy 211 KSGDYAVPRYENLKRLPDPRPFANGQRAVLGWRKRFYRTYDRDRDQGLPAHSHEYD 270
 Db 218 KGADY-----RDGNNNDIDVLLKTHWQLTDSQ-----LAANFHY-- 255

 Qy 271 DCHADIIWQSLINKRYLQLYPHLLTBEEDVDYDNPGLSCGFHDDDDAHAAHNGKPIDL 330
 Db 256 DAYADM-----PGGLTQAQVD--DPPQSV-----RMDNDF 284

 Qy 331 RNKR-----YELRAEWKQPPPGFEALRVHLNR-----NDYHDEKAGDA 369
 Db 285 RGRKDFSLKYTRQVDDLTQFVLTYYSDSFRGSSIAARNLTITTSYPRDYHFAVEPRV 344

 Qy 370 VENFFNQTQNAIRELHOPIGRLKSGWQVYLCKOKSALSATSSEAVKQPMILDN----- 424
 Db 345 SRIFFAFPTT-----QEVG-----IGRYLKE-----AMNERASQALVDNVPTR 385

 Qy 425 -----KVQHSYFFGFEQANWDFNFTLEGVRVEKQKASIRYDKALI----- 464

RESULT 19
 Q9HYX3 PRELIMINARY; PRT; 721 AA.
 AC Q9HYX3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Probable TonB-dependent receptor.
 GN PA3268.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Gotlib L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Lazbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sauer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 RT opportunistic pathogen";
 RT Nature 406:959-964(2000).
 RL EMBL; AE004749; AAC06656.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR InterPro; IPR006195; tRNA_ligase_II.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00862; AA_TRNA_LIGASE_II; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 DR Receptor; Complete proteome.
 SQ SEQUENCE 721 AA; 79371 MW; 70B6716698A9E05D CRC64;

 Query Match 6.1%; Score 246; DB 16; Length 721;
 Best Local Similarity 21.2%; Pred. No. 6e-09;
 Matches 184; Conservative 102; Mismatches 277; Indels 304; Gaps 42;

 Qy 4 TTLKPIVLSILLINTPLLAQAHETEVSGLTETVTVGKSRPRATSGLLHTSTASDKII-S 62
 Db 5 SSLSRLAVALAIGVSMQAQAEEAEKELG--TIVTVGDLGEADQAVVQNHGARSVVR 62

 Qy 63 GDTLRQKAVNLGDALDVGPHGASQYGGGASAPV-----TGTGTRIKVNLNHHGETGDM 117
 Db 63 REMLESQAQNRVDLRKVPQVQVQDNNNGTGGSDISLNVGVRGLTSR----- 108

 Qy 118 ADFSPDHAIWVD-----TALS-----QQVELRGFTVLLYSSGNVAGLVD 157
 Db 109 --LSPRSTWLDGVPAAVAPYQPOLSMPLSIGNLESIDVVRGAGSVRYPQNVGVIN 166

 Qy 158 VADGKIPEKMPENVSSELGLRL---SSGNLEKLT---SGGINIGLGRNFVHLHTEGLYR 210
 Db 167 FVTRAIPEKF-----SGEIGTTIEHAGGKWKLNQAFGLGTADNGLG--VALLYSV-- 217

 Qy 211 KSGDYAVPRYENLKRLPDPRPFANGQRAVLGWRKRFYRTYDRDRDQGLPAHSHEYD 270
 Db 218 KGADY-----RDGNNNDIDVLLKTHWQLTDSQ-----LAANFHY-- 255

 Qy 271 DCHADIIWQSLINKRYLQLYPHLLTBEEDVDYNPGLSCGFHDHDDDAHAHAKGKPIDL 330
 Db 256 DAYADM-----PGGLTQAQVD--DDPFQSV-----RMDNDF 284

 Qy 331 RNKR-----YELRAEWKQPPPGFEALRVHLNR-----NDYHHRKAGDA 369
 Db 285 RGRKDFSLKYTRQVDDLTQFVLTYYSDSFRGSSIAARNLTITTSYPRDYHFAVEPRV 344

 Qy 370 VENFFNQTQNAIRELHQPILGRKSGWQVYLCKOKSALSATSSEAVKQPMILDN----- 424
 Db 345 SRIFFAQPTT-----QEVG-----IGRYLKE-----AMNERASQALVDNVPTR 385

 Qy 425 -----KVQHSYFFGGEQANWDFNFTLEGVRVEKQKASIRYDKALI----- 464

QY 221 --RNKRLPSPRRFANGQRAVLGWRKRYRRTYSDRRDQYGLPAHSHHYDDCHADIW 278
D 222 DANNL-----RLNGGEH-----LTDAYRNQGLAKFGWQADDA-----265
QY 279 QKSLINKRYLQLYPH-----LLTEEDVDYDNPGLSCGPHDDDDAHAAHNG 324
D 266 QRLSLVRQETHQNAFNSQWSTSVPLVQOQTRDF-NTLDRYNPTDSR-----317
QY 325 KPMIDLKRYELRAEWKQPPGPEALRVHLNRNDYHDEKAGDAVEN-----372
D 318 --WLDARVAVNKT-----FDEYRVTKQSDKVDYRTLGLNLSHNSQFDLLSLTYG 368
QY 373 --FNNQOTQARIERHOPI---GRLLGSGWGVQLGKSSALSATSBAVQPMLLDNKVQ 427
D 369 GDYEDYKTOGER-EGKRPPIPADGRSK-VMG-----SYVQADIPLGSGWNLPLG---415
QY 428 HYSFGFGEQANWNTLLEGGVVRVEKQASIRYDKALIDRENYKQPLDPLGAHQRTARSF 487
D 416 -----LRYDHTAE-----DKNIAGSERSEHLSFVSGL-----444
QY 488 ALSGNWFTPOHKLSTASHQE--RLPSTOELYAHGKH-----VATNTFEVGNKHLNKR 540
D 445 ---RWAATDW--LTLNARYDEAFRAFSMEEMYTGTGFCMGPNGCNVFK-ANPDLKPET 497
QY 541 SNNIELALGYE-----GDRQYNLALYRNPFGNYIYAO-----TLNDGRGPKSIE 585
D 498 AKNKEISAQWRFNVLADDELAFATYFHNNSVNIIDQNMFTTQYVNVTDAR-LRGVE 556
QY 586 DDSEMKLVRYN---QSGADPYGAEGEYFKPTPRYRIGVSGDYVRGLKNLP--SLPGR 639
D 557 LDA-----RYNWRDLETSLVAQTEG-----RDKTNQPLNNIPAH 592
QY 640 EDAYGNRPFTAQDONAPRPAARLGHKLKSLTDRIDANLDYRVFAQNKLAARYETRT- 698
D 593 KWLGVSHYFMDRD-----LKAGV-----NVSHVE--SQDELPSNVTVY 630
QY 699 PGHMLNGANRYRNTRYGEMWNVKA-----DNLNQSVAHSSFLSDTPQWGRSFT 751
D 631 PSYTLVDLYTWQPT-----GALKAVKVDVGDINDVIDE--YRQAP-DQLYSAGRNFK 681
QY 752 GGVNVKF 758
D 682 VGVRYSF 688

RESULT 21
Q8EA32 PRELIMINARY; PRT; 708 AA.
AC Q8EA32;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TonB-dependent receptor, putative.
GN S04077.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Sehadi R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis.";

RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AE015838; AAN57051.1; -
DR TIGR; S04077; -
KW Receptor; Complete proteome.
SQ SEQUENCE 708 AA; 78222 MW; 20E5E58517FFAIDF CRC64;
Query Match 6.0%; Score 244; DB 16; Length 708;
Best Local Similarity 20.4%; Pred. No. 8.1e-09;
Matches 171; Conservative 116; Mismatches 312; Indels 240; Gaps 36;
QY 9 IVLSILLINPLAQAHEHETOSVGLTETVTVVVKSRPRATSGLLHTSTASDKIISGDTLRQ 68
D 21 LTLALGLSPAP-LAMANDNEQT-QIERILVHGE-----QSICR 56
QY 69 KAVNLGDALDQVPGIHTASQYGGASAPVIRGQTRRIKVLNHHGETGMDAFSDHAIMV 128
D 57 NALGSADALLKQDQVDFSEAGVSALPVLNGMGDRILKVL-----VDGADITASCANQM 110
QY 129 DTALS-----QOVELRGPVTLILYSSGNVAGLVVDVADGKIPEKMPENGVSGBGLURLS 181
D 111 NPPLSYVSANQIINSVEVAGVSPVSAGGDIAGVIKV-NSLNPKFTDSENLSFESG-EIS 168
QY 182 SGNLEKLTSGGINIG---LGNFVLHTEGLYKSGDYAVPRYRNLKRLPDSPPRANG 236
D 169 SG--YRSTSDSLIGAKAGIASKNVLSYOGAVEDANSY---HDGNGDKVLDITLYRAQN- 222
QY 237 QHRAVLGWRKRYRRTYSDRRDQ-----YGLPAHSHHYDDCHADIWQKSLIN 284
D 223 -HALTAWR-----DEKQQLAVLTKHAQIPFGFPNQYMDMTD-----N 260
QY 285 KRYLOLYPHLTHEDVDYDNPGLSCGPHDDDDAHAAH-----NGKPWIDLNRKY 335
D 261 KSYGALVRYLR-----DLENDG---EFSQALNHSVKHEMGFFTPPEKTGKMPNTEGSDY 312
QY 336 ELRAEMKQPPGPEALRVHLNRNDYHDE-----KAGDAVENFNNQTNQARIELR 386
D 313 SYQLHRLTWGDDSTLLGQEYYSYQLDDIWPVPGTMAPNDVININDGERRAAV--- 369
QY 387 HQPIGRKLSGWGVQLGKSSALSATSBAVQPMLLDN--KVQHSYFFGVEQANWNTFL 444
D 370 -----YGEW-----QQNLNPLWLSAGVRYEVTTDTGCVQAYS-----NWPM 407
QY 445 EG--GVTVKQKASIRYDKA---LIDRENYKQPLDPLGAHQRTARSFALSGNWTTPQ 498
D 408 MGMPNVDAEAAKAFNAMDRSRDNLIDA-----TLIARYQLSAK 446
QY 499 HKLSLTASHOERLPSTOELYAHGKHVATNPF-----EVGNKHLNKRSSNNIELALG 549
D 447 QQLFGLARKNRPANLYERSWGRGVWATTWI GWYGDGNGYVGNPDLKPETAHTLSAYK 506
QY 550 YEGDRWQYNLALYRNPFGNYIYAQTLNDGRGPKSIEDDSMKLVRYNQSGADFYGARGEI 609
D 507 FNGDAWQFSTAWYSAVTDYDAEVI--GSFNRTSTPAKRNILKLTNEDANLYGAKFSA 564
QY 610 YEKETPRYRIGVSGDYVRGLKNLP--PQREDAVGNRPFAQ-----DD 653
D 565 LY-----LLADTDSGKQWMLKLNITRGERDEGNEPLYQIKPLQTLALSHQLGDW 615
QY 654 QNAPRVPAARLGFHLKASLTDRIDANLDYRVFAQNKLAARYETRTPGHMLNLCANRYRN 713
D 616 EN-----RLAQWVAT-KDRVDDR-----RLENETSSYLLNLS-----S 650
QY 714 TRYGENWYVKADNLLNQSVYA-----HSSFLSDTPQWGRSFTQGVNVKF 758
D 651 IKMQELSLTAPAINNLF-DTYVELPLGGSVIAEFKADSSNGFSQVAGSGRSFELGASVRF 708

RESULT 22
O66812 PRELIMINARY; PRT; 680 AA.
ID O66812
AC O66812;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)


```

QY 497 POHKLSTASHOERLPSTQELYAHGKH---VATNTFEVGNKHLNKNERSNNIELALGYEGD 553
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 465 --HQLYFSYNHGVRAPYDKAYGVSHEDIPFPFIIPNMDLKAETSDSFEVGHKFDNG 522
QY 554 RWQYNLYALYRNRRGNYYIAQTLDNCRGPKSIEDSEMKLVRY-NOSGADFYGABGEIYFK 612
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 TSLYLYVALFYQQDDFI-----DVKEVGMVGPVTLKQYQNLNGVETYGAELSAAHQ 573
QY 613 PTPRYRIGVSGDYVRGR-----LKNLPSPFGR-EDAYGNRPFFIAQDDQNAQPRVPAARLG 665
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 574 INEQMLLATKAGYVDGKNDQGEYIRTLTPVEGNVELSYAKQNW-----SASLL 621
QY 666 FHLKASLUTRIDANLDYRVFAQNKLYE---TRTPGHMLNLGANYRNRTRYGEWNWY 722
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 WNWAASM-DRVPVCVNDF-----NQEVPCAKTESWHSFDFLASY-----QWN-- 662
QY 723 VKADNLLNQSV-----YAHSSFLSDTPQMGRSFTGGVNVKF 758
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 663 --KDLTUSASIVNLFNDEYRYQDVAGISQAHTTY---STEPGRYFTANVKYQF 711

RESULT 24
O68881 PRELIMINARY; PRT; 764 AA.
AC O68881;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane hemin receptor.
GN PHUR.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PAOI;
RX MEDLINE=96210657; PubMed=8633080;
RA Ochser U.A., Vasil M.L.;
RT "Gene repression by the ferric uptake regulator in Pseudomonas
RT aeruginosa: cycle selection of iron-regulated genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4409-4414 (1996).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=PAOI;
RA Ochser U.A., Vasil A.I., Johnson Z., Vasil M.L.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL: AF055999; AAC13289.1; -.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Rec; 1.
KW Membrane; Outer membrane; Receptor; TonB box.
SQ SEQUENCE 764 AA; 84650 MW; 619C71FC49D00D41 CRC64;

Query Match 5.9%; Score 237.5; DB 2; Length 764;
Best Local Similarity 20.0%; Pred. No. 2.7e-08;
Matches 173; Conservative 125; Mismatches 328; Indels 241; Gaps 35;

QY 6 LKPIVLSILIN-----TPLLAQAHETEQSGVLETVTVVGKSPRATSGILLHT 53
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 LRP-CLALLLSPLSALAGNAVPLTPTTATRTQEAQVDSVPSTVSQTRQLD----- 61
QY 54 STASDKLIISGDTLRQKAVNLGDALDGVGGHAGSYGGASAPV-----IRGQTRRIKVLN 109
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 -----KQNVNNIKELVRPEPV--SVGGAGQAGAGITGVNIRGIDGNRIILQI 106
QY 110 HHGETGD---WADPSDPHAIWDTALSQQVEILRGPVTLTSSGNVAGLVD----- 157
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 DGVLPNDPFGSGYQATHRYNVPDIVKRVIELRGPASALYGSNAGVSYFTLPDSI 166
QY 158 VADGKIPKMPENGVSSELRLSSGN-----LEKLTSGG-----INIGLGNFV 202
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 167 IKDGK-----DVGARLKAGYESASHSWLTSATVGRADDGLGLHYGRQCHE 214
QY 203 LHTEGLYRKSGDVAVPRYRNKLKLPDPSRRFANGQHEAVLGWR-----KRPYRR 251
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 TESNGGHGGTG-----LSRSEANPEDADSYSLKGLQWYAEGRGLVFEKYKSD 265
QY 252 TYSRRDQYGLPAHSHEYDDCHADI-----IWQK---SLINKRYLYQVPHLITEE 298
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 VDTDQKSAYGGP-----YDKGKPAIPDPMPLPGGMYQWRKGNDAITRERYGLEHHFLDSQ 320
QY 299 DVDYDNPGLSCGPHDDDDAHAAHNGKPMWDLNKNRYELRAEWKPPPGPEALRVHLNRN 358
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 VADRIQWSLYAQAKTDQATREFY-----YPTTRKVIJTR 355
QY 359 DYHDEKA-----GDVAFENFNNQTONA---RIELRHQPIQRLKSGMWQVYLGQKSSALSA 411
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 DTTYKERLWVFDSQLDKSFAIGTEHLLSVGINLKHQKVTGMRSGTGNLDTGADSPDA 415
QY 412 TSAVQKQPMLLDNKVQHSYFFGVEQANWNTFLEGGVVE-----KOKA 455
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 416 LERSSDPP---DPTVKTYALFAQDSISWNDWTFPTGLRYDYTRMEPHITDBFLRTMQSQ 472
QY 456 SIRYDKALIDRENYKQPLDGLGHRQTARSPALSGNNWYFTPOHKLSTLASHOERLPSTQ 515
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 473 NTADES-----DKKWHRVSPKFG---VTYDFAQHYTWY--GQYAAQGF-----RTPTAK 516
QY 516 ELYAHGKHVATNTFEVGNKHLNKNERSNNIELALGYEGDRWQYNLYALYRNRRFGNYIAQTL 575
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 ALYGRFENLQAGYHIRPNPNLKPESQSFETLGRGKFDGSGFVAVFYNKYRDPIDEDAL 576
QY 576 N-----DORGPKSIEDDSEMKLVYNSGADF-----YGARGEIYFKTPPYRIGVSG 623
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 577 NTDSGTGGNGQTFQSNNIERAVIK---GVELKGRLELGAFAQGLYTOGSVAYAYGRNK 632
QY 624 DYVRGR-LKNLPSLP-----GREDAYGNRPFFIAQDDQNAQPRVPAARLGFHLKASLTDRID 677
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 633 D--NGEPINSVNEPLTGVFLGYDEADGN-----YGGLLSWTLVLRKRD 672
QY 678 ANLDYRVFAQNKLARYETTPGHMLNLGANYR-----RNTRYGEMWYV 723
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 673 -RVDDSTFHAPDGTAS-QFKTPGFGVLDLSAYYRLSKDLTLNAGLYNLTKKYWLWDVVR 730
QY 724 KADNLLNQSVYAHSSFLSDTPQMGRSF 750
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 GYDSVGEASALAPAN-IDRLSQPGRNF 756

RESULT 25
O9X5P3 PRELIMINARY; PRT; 756 AA.
AC O9X5P3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane heme receptor PfHR.
GN PHHR.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=ATCC 15453;
RA MEDLINE=20121752; PubMed=10658665;
RA Ochser U.A., Johnson Z., Vasil M.L.;
RT "Genetics and regulation of two distinct haem-uptake systems, phu and
RT has, in Pseudomonas aeruginosa.";
RL Microbiology 146:185-198 (2000).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL: AF127222; AAD31012.1; -.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_dep_Rec; 1.
KW Membrane; Outer membrane; Receptor; TonB box.

```

SQ	SEQUENCE	756 AA; 83202 MW; B499B739D04E25D5 CRC64;
DT	Query Match	5.8%; Score 235.5; DB 2; Length 756;
DN	Best Local Similarity	20.9%; Pred. No. 3.7e-08;
GN	Matches 176; Conservative 119; Mismatches 336; Indels 213; Gaps 37;	
OC	20 LLAQAHETEQSGVLETVTVVVGKSRPRATSGLLHTSTAGDKIISGDTLRQKAVNLGDALDG 79	
OS	15 IVLEKRSAGSALAQQVTSATRNEQDYNVPSTVSVHDR---AELDRQHVNIRELVRY 71	
OC	80 VPGIHASQYG--GGASAPVIRGQTGRRIRKVLNHHGETGD---MADFPDPAIMVDLTALSQ 134	
OS	72 EPGSVGGAGTRSGNAGYINRIGIDGRVLTQDVGVEPDNPFNGFYAKTRRNYVDPEIVK 131	
OC	135 QVEILLRGPVTLIYSGNVAGLVD--VADGKPEKM--PENGVSGBELGLRLSGNLEKLT 190	
OS	132 RVEILLRGPVTLIYSGNVAGLVD--VADGKPEKM--PENGVSGBELGLRLSGNLEKLT 188	
OC	191 GGINIGLKNF--VLHTEGLYRKSG---DYAVPRYRLKELPSPRRFANGQHRVILGW 244	
OS	189 GTF-AGRVQDFGLH---LSQRNGHETESYDGNATGLARTGANPED-ARTNVLAKLW 243	
OC	245 RKRFRYRTYSRRQYGLPAHSHYEDDCHADIIWOKSLINKRYLQLYPHLLTEEDVDYDN 304	
OS	244 -----NYG-----DDNRLGLTYEK----- 257	
OC	305 PGLSCGFHDDDAHAHANGKPMI-----DLNKEYELRAEWKQPPGCFE 349	
OS	258 -----FKDDRVLNKNVAGPFGGRRGNLYRRRRGNDTITRERFGLNTEFALESPIAD 311	
OC	350 ALRVHLNRNDYHDE-----KAGDAV-----ENFFNQTONA----- 381	
OS	312 RIKTSLVNQIAKTDQTTAEIYQAGRRVLTRDTLYEEKQWFDQADLKAFSLGETDQV 371	
OC	382 -RIELRHOPI-GRUKSGWQVYLOKSSALSA---TSEAVKQPM-LDNKVQHYSPFGV 434	
OS	372 YGTTLKQKQVTSREGSASCLAIGAAGCTAIGAPSPASDSVKASDFDPDTINTYSLFAQ 431	
OC	435 EQANWDFTLBGGVVRVEKQASIRYDKALIDREN---YKQPLDPLGAHROTAR---SFA 488	
OS	432 DQITWDKWTFIPAVRYDYTRLKPQLTQEFNLVNPQTGAYTVGDKQKWNVRVTPKFLTYA 491	
OC	489 LSGNWYFTPQKLSUTASHQRLPSTQELYAHGHVATNTFEVGNKHLNKNERSNNIELAL 548	
OS	492 LTDNVTFGQY-----AEGFRTPSAKALYGRFENLNGYTVENPDLPKPTSKGIEGI 545	
OC	549 GYEGRWQYNIALYRNRPGVI-----YAQTLNDGRGP-KSIEDDSEKMLVRY 595	
OS	546 RGKFDGSDFIADVYNYKYRDFIDEDKPVAGGTVEQFQAVNIKRATIKGVEAKGRNL--- 602	
OC	596 NQSGADFYGABGEIYFKPTPRYRIGVSGDYVRGR-----LKNLPSPGREDAYGNR-P 647	
OS	603 -----DTLGAPKGLYTQGSVAYTYGRNDD--NGEFLNSVNLKGVFGYQDDNYGVLVS 655	
OC	648 FIAQDDQNAIPVPAARLGLFHLKASLTDRIDANLDLYRVFAQNKLARVETRTFGHMLNLG 707	
OS	656 WTVVKQN--RVDSTT--FHAPDGGTDG-----PFKTPGFGILDLT 692	
OC	708 ANYR-----RNRTRYGEWNW-YKADNLLNQSVTAHSSFLSDTPQMGSRFTG 752	
OS	693 AYYKVSQDVTNGLYNLTDKKY--WNWDDVRSYDSVGEAGVTGPANLDRLTQPGRNF-- 748	
OC	753 GNVN 756	
OS	749 AINV 752	
OC	PRELIMINARY; PRT; 764 AA.	
OC	Q9HV88	
AC	Q9HV88	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	

DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DN	Probable outer membrane hemin receptor.	
GN	PA4710.	
OC	Pseudomonas aeruginosa.	
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;	
OC	Pseudomonadaceae; Pseudomonas.	
OX	NCBI_TaxID=287;	
RN	[1] SEQUENCE FROM N.A.	
RP	STRAIN=ATCC 15692 / PAO1;	
RC	MEDLINE=20437337; PubMed=10984043;	
RX	Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,	
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,	
RA	Garber R.L., Coltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,	
RA	Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,	
RA	Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,	
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;	
RT	"Complete genome sequence of Pseudomonas aeruginosa PAO1, an	
RT	opportunistic pathogen."	
RL	Nature 406:959-964(2000).	
DR	EMBL; AE004885; AAG08096.1; -.	
DR	InterPro; IPR000531; TonB_box.	
DR	Pfam; PF00593; TonB_dep_Rec; 1.	
KW	Receptor; Complete Proteome.	
SQ	SEQUENCE 764 AA; 84724 MW; EA45897F2334255A CRC64;	
Query Match	5.7%; Score 231.5; DB 16; Length 764;	
Best Local Similarity	19.8%; Pred. No. 7.3e-08;	
Matches 172; Conservative 126; Mismatches 325; Indels 247; Gaps 35;		
QY	6 LKPIVLISLLIN-----TPLLAQAHETEVSGLTETVTVVGKSRPRATSGLLHT 53	
DB	9 LRP-CLALLLSPSLALAGNAVPLTPTTITRTEQAQVDSVPSTVQTRQLD----- 61	
QY	54 STASDKIISGDTLRQKAVNLGDALDVGPIHASQVGGASAPV---IRGOTGRIRKVLN 109	
DB	62 -----RQNVNNIKELVRYEFGV--SVGAGQAGITGTYNRIGDGNRLTQI 106	
QY	110 HHGETG---MADFPDPAIMVDLTALSQQVILRGPVTLIYSGNVAGLVD----- 157	
DB	107 DGVLEPNDFSGVAQTHRNVDPIVKRVEILRGPSALYGSNAIGGAVSYFTLDPSDI 166	
QY	158 VADGKIPEKMPENGVSGBELGLRLSSG-----NLEKLT-----GGINIGLKNPV 202	
DB	167 IKDKK-----DVGARLKAGYESASHSLTTSATVAGRADDGDLHYGRQGE 214	
QY	203 LHTEGLYRKSGDYAVPRYRLKRLPDSPPRFANGQHRVILGW-----KRFYRR 251	
DB	215 TESNGHGGTG-----LSRSEANPEDADSYLLGKLGWYAEGRFGLVFEKYSQ 265	
QY	252 TYSRRDOYGLPAHSHYDDCHADI-----IWOK-----SLINKRYLQLYPHLLTEE 298	
DB	266 VDTDQKSAYGFP-----YDKGKPAIPPSMLPGMYQWRKGNLTLTRERYGLEHFLDLSQ 320	
QY	299 DVDYDNFGLSCGFHDDDDAHAAHNGKPIWDLNKRVELRAEWKQPPFGPALRVHLNRN 358	
DB	321 VADRIQSLNYQLAKTQATREFY-----YPTIRKVLRT 355	
QY	359 DYHDEKA---GDAVENFFNNQTQNA---RIELRHOPIGRLKGSWGQYVLGQKSSALSA 411	
DB	356 DTTYKERLWVFDSDQDKSFAIGETHELLSYGINLKHQKVQWGRSGTGNTLDGADSPDA 415	
QY	412 TSEAVKQPMLLDNKVQHYSPFGVEQANWDFLTBGGVRVE-----KQKA 455	
DB	416 LERSDDFP---DPTVKTYALFAQDSISWNWDVTFPLGRYDTRMEPHITDFLRTMKQSQ 472	
QY	456 SIRYDKALIDRENYKQPLDPLGAHQRTARSFALSGNWYFTPQKLSLTASHQRLPSTQ 515	
DB	473 NTAVDSE---DKKWHRVSPKFG---VTYDFAOHTYTWY--GQYAQGF-----RTPTAK 516	
QY	516 ELVAGKHVATNTFEVGNKHLNKNERSNNIELALGYEGDRWQYNLALYRNFENGYIYQTL 575	
DB	517 ALYGRFENLQAGYHIEFNPNLKPEKSSQSFETGLRGKFDGSGFVAVFYNKTRDFIDEDAL 576	

```

QY 576 N-----DGRGPKSIEDSEMKLVRYNQSGADF-----YGAEGRIYFKPTPRYRIGVSG 623
Db 577 NTDSTGNGQOTQSNRIERAVIK-----GVELKGRLELGAFAQOGLYTGQSVAYAYGRNK 632
QY 624 DVVRGR-LKNLSLP-----GREDAYGNRPFTAQDDQNAAPRVPAAELGFLKASLT---D 674
Db 633 D--NGEPINSNVPLTGVFLGVDYDEADGN-----YGLLSWTILVKRKD 672
QY 675 RIDANLDYRVFAQNKLARYETRTPGHMLNLGANYR-----RNTRYGEWN 720
Db 673 RVDDS-----THTPDGTASQKFTGFGVLDLSAYYRLSKDLTLNAGLYNLTDKKYWLWD 727
QY 721 WYVKADNLNQSVYAHSSFLSTPQWGRSF 750
Db 728 DVRGYDSVGEASALAPAN-IDRLSQPGRNF 756

RESULT 27
Q8ZG14 PRELIMINARY; PRT; 665 AA.
AC Q8ZG14;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Putative outer membrane protein (Putative outer membrane
iron/siderophore receptor).
GN YP01313 OR Y1IUR OR Y2872.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
[1]
[2]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Meule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
[2]
SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL: AJ414147; CAC90143.1; -
DR EMBL: AE013889; AAM86423.1; -
DR InterPro: IPR000531; TonB_boxC.
DR Ffam; PF00593; TonB_dep Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC; 1.
KW Hypothetical protein; Receptor; Complete proteome.
SQ SEQUENCE 665 AA; 73909 MW; F5ACA23F30B06871 CRC64;

Query Match
Best Local Similarity 5.6%; Score 228, DB 16; Length 665;
Matches 152; Conservative 116; Mismatches 280; Indels 188; Gaps 34;

QY 24 AHETQSGLVETVTVVVGKSRPRATSGL---LHTSTASDKIIISGDTLRQKAV-NLGDALDG 79
Db 29 AEKTYATPTDMV-----TAGGFOQRIQDSAAISVTVREQIENKAYRDVTDALKD 81
QY 80 VPGIHASQYGGASAPV-IRGTGRRIKVLNHHGTGMDAFSP-----DHAIWDTA 131

```

```

Db 82 VEGVWIT--GGSTSDISIRGMAAKYTLIL-VNGKRVDTTRTPNSDGGSGIQGMLPPLA 138
QY 132 LSQQVEILLRGPVTLTYSSGNVAGLVADVADGKIPEK-----MPENGVSSELGLRLS 181
Db 139 AIDREIVRVGPMSSLYGSDMGGVINIITRKVGKEWHGTVRADATLQEDSKSGDI----- 193
QY 182 SGNLEKLSGGGINIGLGRNFVLHTEGLY-RKSGDYAVPRYRLKRLPDSRPFANGQRA 240
Db 194 -POTNAYASGPLIDGL--LGLKVSGLLSHRSEDKIIDGY-----NQRMNRGTATF 241
QY 241 VLGRWKRFRYRRYSDRDOYGLPAHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTEEDV 300
Db 242 TL-----TPDDNNEFFDIGHYVQDRNSTPQRTLALNG---TNSDT 279
QY 301 DYDNPGLSCGFHDDDDAHAHNGKFWIDLRNKRVELRAEWKQPPFGFEALRVHLNRNDY 360
Db 280 QYDR-----NNYAVTHNG--YYDFGNS-----TSYIQDET 308
QY 361 HDEKAGDAVENFFNNTQNAEIRHOPIGRLKSGWGVQYLGOKSSALSATSEAVKQPM 420
Db 309 RNPSRQMKSVDNIFN--TQTSFLDNHTLI--LGGQYRYEELYDKGNQLPSASDL----- 359
QY 421 LLDNKKVQHSYSPFGEQANW---DNFTLGGVREVEKASIRYDKALIDRENYKQPLPDL 477
Db 360 ---KKLTRWSMALFAEDBQWMTNDFTLGGIRMDQD-----ONY----- 395
QY 478 GAHQRTARSFALSNNWYFTTPQHKLSLTASHQRLPSTQ-----ELYAHGKHVATNTE 530
Db 396 GTH-WTPR--LYGVWHLADQWTLKGGVGGYRSPDLRQATDDWGQSGGKG-GLPALI 450
QY 531 VGNKHLNKNERSNNIELALGYEG-DRWQYNLALYRNRPNGYI-----YAQTND 577
Db 451 LGNSNLKPERSTISQIEGILMDQEGWNASVTLYTFDFDKKITEVRNCDITTTNTGQCFN 510
QY 578 GRGPKSIEDSEMKLVRYNQSGADFYGAEGRIYFKPTPRYRIGVSGDYVRGLKNLPSLP 637
Db 511 GINYKFISSD-----RINVDKAMTRGAEATPAWDINQAWSLATNYTTFQSSQKS----- 558
QY 638 GREDAYGNRPFTAQDDQNAAPRVPAAELGFLKASLTDRIDANLDYRVFAQNKLARYE-- 695
Db 559 ---GAFAQPL-----NQMPKMLNGLTNWKTTFEDATWIRAN---YRGKASEVLENTSMG 608
QY 696 TRTPGHMLNLGANYR 711
Db 609 SRTPSYTFVDLGANYQ 624

RESULT 28
P72121
ID P72121 PRELIMINARY; PRT; 723 AA.
AC P72121;
DT 01-FEB-1997 (TremBLrel. 02, Created)
DT 01-FEB-1997 (TremBLrel. 02, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Outer membrane protein C precursor (Outer membrane protein OPRC).
GN OPRC OR PA3790.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
[1]
[2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=96349120; PubMed=8760927;
RA Yoneyama H., Nakae T.;
RT "Protein C (OprC) of the outer membrane of Pseudomonas aeruginosa is a
copper-regulated channel protein.";
RL Microbiology 142:2137-2144(1996).
[2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;

```

RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P., Hickey M.J., Brinkman F.S.I., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Labbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.; "Complete genome sequence of <i>Pseudomonas aeruginosa</i> PAO1, an opportunistic pathogen.";	DB	5.6k; Score 224; DB 16; Length 723; Best Local Similarity 19.4k; Pred No. 2.3e-07; Matches 165; Conservative 129; Mismatches 281; Indels 276; Gaps 39;	QY	13	ILLINTELLAQAHTEQSVGLTETVTVGKSPRATSGLLHTSTASDKIISGDTLRLQAVN	DB	44	LALSPSALLAEHSQHODHAVELA-----PSVVTGVAQSSPLT-IVTNPKPRQVP	QY	73	LGDAID--GVPGIHAQYGGASAPVIRGTGTRIKVLNHHGETGMADSPHAINVD	DB	95	ASGADYKLTIPGPAVIRNGSGNDPVLRMFGSRMLITNGM--MLGACPRNMDAPT	QY	130	TALS-----QQVEILRGPVTLIYSSGNVAGLVADVADGKIPEKMPNGVSGELGLRLSSGNL	DB	152	SYISPETYDKLTIVKGPQTVLWGPGASAGTI-----LFEPEPER--PGELGSRVNASLL	QY	186	EKLTSGGINTGLGNKFNVLHTEGLYRKSGDYAVPYRNKLUPDSPRRPANGQHRVILGWR	DB	204	-----AGSN-----GRFDKVLDD-----AAAGNR	QY	246	KRFVRRYTSRRDQYGLPAHSHEYDDCHADII-----WOKSLINKRYLQLYPHLLTEEDVDY	DB	222	LGVLRT-----GNHAQSDDEDGAGNVTPSRWKKN-----	QY	303	DNPLSCGPFHDDDAHAHAHNGKPMIDILRNKRYELR-----AEMKQPFPGFALRVHLNRN	DB	253	-NGDVAVGWTPDEDTLIELTAGK--GDGEARYAGRGMDGSQFKRESLGLRFVKSVNS--	QY	359	DYHDERKAGDAVE-----NFFNNQTQNAARI-----ELRHOPI-----GELK	DB	307	-----DVLEKVEAQVYNYADHIMDNFRLTPTDPPSSMMPMPMASQVDRRTLQGERLA	QY	395	GSW-----GVQYILGQKSSALSAT-----SEAVKQPMLLDNKVQHSYFPGVEQA	DB	358	ATWRWDDFKLVGTGDAMRNEHRARGSKYDMMTDYITDADQFPWSKDAVFHNYGAFG--EL	QY	438	NW-----DNFTLEGVYRVKQKASIRYDKALIDRENYIKQPLDPLGAHQRTARSFALSGNWY	DB	416	TWFAAERDRLLIGGLRLD--RASVK-----DYRQTLKSGHM-GHAM	QY	495	FTP-----QHKLSLTAS-----HOERLPSTQELVA--HGKHVATNT	DB	453	ANPTANDTRADTLPSGFVRYEHLADSPFTLYAGIGHAERPDYWEILFSPKRGPNGSVNA	QY	529	FEVGNKHLNERSNNIELALGYEGDRQWYNALYRNFRGNVYIAQTLNDGRGPKSIEDDS	DB	513	FD-----KIKPEKTTQLDQFLQYNGDKLQAWASGVYGVVQDFILFSYREGMMGSSQTATNV	QY	589	EMKLVRYNQSGADPYGAEGEITYFKPTPRYIRGVSGSDYVRGELKNLPSLPGREDAYGNRP	DB	569	DARIM-----GSELGASYQLTGNWKTDASLAYAWG--KN-----	QY	649	IAQDDQNAVPAAARLGFHLKASLITRIDANLDYIRYFA-QNKLARVE-----TRTP
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P., Hickey M.J., Brinkman F.S.I., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Labbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.; "Complete genome sequence of <i>Pseudomonas aeruginosa</i> PAO1, an opportunistic pathogen.";	DB	5.6k; Score 224; DB 16; Length 723; Best Local Similarity 19.4k; Pred No. 2.3e-07; Matches 165; Conservative 129; Mismatches 281; Indels 276; Gaps 39;	QY	13	ILLINTELLAQAHTEQSVGLTETVTVGKSPRATSGLLHTSTASDKIISGDTLRLQAVN	DB	44	LALSPSALLAEHSQHODHAVELA-----PSVVTGVAQSSPLT-IVTNPKPRQVP	QY	73	LGDAID--GVPGIHAQYGGASAPVIRGTGTRIKVLNHHGETGMADSPHAINVD	DB	95	ASGADYKLTIPGPAVIRNGSGNDPVLRMFGSRMLITNGM--MLGACPRNMDAPT	QY	130	TALS-----QQVEILRGPVTLIYSSGNVAGLVADVADGKIPEKMPNGVSGELGLRLSSGNL	DB	152	SYISPETYDKLTIVKGPQTVLWGPGASAGTI-----LFEPEPER--PGELGSRVNASLL	QY	186	EKLTSGGINTGLGNKFNVLHTEGLYRKSGDYAVPYRNKLUPDSPRRPANGQHRVILGWR	DB	204	-----AGSN-----GRFDKVLDD-----AAAGNR	QY	246	KRFVRRYTSRRDQYGLPAHSHEYDDCHADII-----WOKSLINKRYLQLYPHLLTEEDVDY	DB	222	LGVLRT-----GNHAQSDDEDGAGNVTPSRWKKN-----	QY	303	DNPLSCGPFHDDDAHAHAHNGKPMIDILRNKRYELR-----AEMKQPFPGFALRVHLNRN	DB	253	-NGDVAVGWTPDEDTLIELTAGK--GDGEARYAGRGMDGSQFKRESLGLRFVKSVNS--	QY	359	DYHDERKAGDAVE-----NFFNNQTQNAARI-----ELRHOPI-----GELK	DB	307	-----DVLEKVEAQVYNYADHIMDNFRLTPTDPPSSMMPMPMASQVDRRTLQGERLA	QY	395	GSW-----GVQYILGQKSSALSAT-----SEAVKQPMLLDNKVQHSYFPGVEQA	DB	358	ATWRWDDFKLVGTGDAMRNEHRARGSKYDMMTDYITDADQFPWSKDAVFHNYGAFG--EL	QY	438	NW-----DNFTLEGVYRVKQKASIRYDKALIDRENYIKQPLDPLGAHQRTARSFALSGNWY	DB	416	TWFAAERDRLLIGGLRLD--RASVK-----DYRQTLKSGHM-GHAM	QY	495	FTP-----QHKLSLTAS-----HOERLPSTQELVA--HGKHVATNT	DB	453	ANPTANDTRADTLPSGFVRYEHLADSPFTLYAGIGHAERPDYWEILFSPKRGPNGSVNA	QY	529	FEVGNKHLNERSNNIELALGYEGDRQWYNALYRNFRGNVYIAQTLNDGRGPKSIEDDS	DB	513	FD-----KIKPEKTTQLDQFLQYNGDKLQAWASGVYGVVQDFILFSYREGMMGSSQTATNV	QY	589	EMKLVRYNQSGADPYGAEGEITYFKPTPRYIRGVSGSDYVRGELKNLPSLPGREDAYGNRP	DB	569	DARIM-----GSELGASYQLTGNWKTDASLAYAWG--KN-----	QY	649	IAQDDQNAVPAAARLGFHLKASLITRIDANLDYIRYFA-QNKLARVE-----TRTP

Db	601	-SSDRALPQIPPLEARFGLTYE	EGDWSAGSL--WRVVAPQNR	IARDQGNVV	GKDFDKSA	657	
Qy	700	GHHMLNLGANYR--	LNTRYGEMWYV	KADNLINOSV	YAHSSPLSD-----TPQMG	747	
Db	658	GFGVFSLNGAYR	VRTRNVKLS-----	AGVDNLF	DKDYTEHLN	KAGDAGFCFSANETVPEPG	712
Qy	748	RSFTGGUNVKF	758				
Db	713	RTEWTKVDFSF	723				
RESULT 29							
Q9S8B9							
ID	Q9S8B9	PRELIMINARY;	PRT;	713	AA.		
AC	Q9S8B9;						
DT	01-MAY-2000	(TREMblurel. 13, Created)					
DT	01-MAY-2000	(TREMblurel. 13, Last sequence update)					
DT	01-MAR-2003	(TREMblurel. 23, Last annotation update)					
DE	Ferric enterobactin receptor.						
GN	FETA.						
OS	Neisseria gonorrhoeae.						
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;						
OC	Neisseriaceae; Neisseria.						
OX	NCBI_TaxID=485;						
EN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=FA1090;						
RK	MEDLINE=99235773; PubMed=10217784;						
RA	Carson S.D., Klebba P.E., Newton S.M., Sparling P.F.;						
RT	"Ferric enterobactin binding and utilization by Neisseria						
RL	gonorrhoeae", 181:2895-2901(1999).						
J. Bacteriol.	181:2895-2901(1999).						
CC	-1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).						
DR	EMBL; AF115385; AAD29612.1; -						
DR	InterPro; IPR000531; TonB box.						
DR	Pfam; PF00593; TonB dep Rec; 1.						
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.						
KW	Membrane; Outer membrane; Receptor; TonB box.						
SQ	SEQUENCE 713 AA; 78864 MW; F24D556E383C90FB CRC64;						
Query Match							
Best Local Similarity 21.6%; Score 216; DB 2; Length 713;							
Matches 187; Conservative 103; Mismatches 298; Indels 278; Gaps 47							
Qy	11	LSILLINTPL-LAQAHTEQSV	GLTVTVVQKSRPRATSGLLHTSTASDKIISGDTLRQK	69			
Db	8	LSLSLTLAAGFAHAENAN	VALDTVTVKGDQ-----GSKRTNIV---TLQOK	55			
Qy	70	---AVNLGDALDGVGHASQYGGGASAPVI	-----RGQTGRRIKY-----LNHH	111			
Db	56	DESTATDMRELLKEEPTI	---DFGGNGTSTFLLTRGNGQNSVDIKVDNAYS	SDSQILYHQ	112		
Qy	112	GETGDMADSPDHAIW	VDTLASQOVELIRGPVTLTLYSSGNVAGLV	VDVADGKIPKEMPE	--169		
Db	113	GR-----FIVDPALVKVSVQKG	-----AGSASAGIATNGALIAKTVAQ	153			
Qy	170	---NGVSGELGRUSSNLEKLTSGGINIG	---LCK--NFVLHTEGLY-----RKSGDYAV	217			
Db	154	DLKLGLDKNGVRLNSG---	FAGNNGVSYGASVFEGKGNF-----DGLFSYNRNDEKDEYA	206			
Qy	218	PR-YRNL---KRLPDSP-----	RRFANGQHRVILGWRKRFYR--RTYSRRDQY	260			
Db	207	GKGFERNVNGKTVPYSA	LDKRSYLAKITTPGQDGHRIVLSHMKDQHRGIRTV	---REF	263		
Qy	261	GLPAHSHEDDCHADII	WQKSLINKRYLQLYPHLLLTEDVDYDNPGLSGCGFHDDDDAH	320			
Db	264	AVGGEN-----SRITIKRQAPAYR	---TTSQNTNLAYTKDGLGFVEKLDANAY	309			
Qy	321	AHNGKPIWIDLRNKEYE	LRAE---WKQFPQPEALRVHLNRNDYHHHDEKAGDAVENFFNQ	377			
Db	310	V-----LEKRYISADDK	NDNGVAGNVKGNPHNTRIAFGNPNFDSRLAE-----Q	353			


```

QY 378 T-QNARIELRHO---PIGRKLGSGVQYLGQKSSAL-----SATSEAVKQPMLLDNKVQ 427
Db 354 TLKYGINVRHQIKPOAFNLNQFKIE---DKDQATEEDKKKARENKIAKAYRLNPTK 410
QY 428 HYSFFGVEQAN-WDNFTLEGGVVEKOKASIRYDKALIDRENYKQPLDGLGAHOTARS 486
Db 411 TDTGAYIEATHEIDGFTLTGLAYDRPKVTHDGK-----TVSS 449
QY 487 FALSGNW--YFTQHKLSLTASHQ--ERLPSTOE-LYAHGKH-----VATNTEFVGNKHLN 537
Db 450 SSINPFGVWQPREHWSFGSNHYSRSPRLDALQTHGKGIISTADGT-----K 501
QY 538 KERSNNIELALGYEGDRWQYNLALYR-----NREFGNIYAQTLNDG-----578
Db 502 AERARNTIEGFNYNDGTFAANGSVFROTIXDALANPQRHDSVAVRANAGVIKONGYE 561
QY 579 -----RGPKSIEDSEMVKLYRNOSGADFYGAE-GEIY-----FKP 613
Db 562 LGASYRTGGLTAKVGVSRSPRFVDTHPKLLSANPE---FGAQTGRTWTASLAYREFK- 616
QY 614 TPRYRIGVSGDYVRGLKNLPSLPGRDAGVGNRPFTAQDDONAPRVPAARLGFHLKASLT 673
Db 617 NPMLEIGWRGYY-----QKATGSILAAGQKDRDGKLENVVRQGFVN-----659
QY 674 DRIDANLDYRVEAQNKLARYETRTPGHMLNLGANYRRNTRYGEMWYVKADNLLNOSV 733
Db 660 -----DVFANMKPLGKDTLN-----VNLNVN-----NVPDKFY 687
QY 734 YAHSSFLSDT-PQMGSRFTGGVNVKF 758
Db 688 YPHSQRTWNTLPGVGRDVRLLGVNYKF 713

RESULT 30
Q51162
ID Q51162 PRELIMINARY; PRT; 720 AA.
AC Q51162;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Re-regulated protein B precursor.
GN FRPB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H44/76; PubMed=7558339;
RX MEDLINE=96009786; PubMed=7558339;
RA Fettersson A.M., Maas, van Massenaar, van der Ley, Tommassen;
RT "Molecular characterization of FrpB, the 70-kilodalton iron-regulated
RT outer membrane protein of Neisseria meningitidis.";
RL Infect. Immun. 63:4181-4184(1995)
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; X89755; CAA61902.1; -.
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22 POTENTIAL.
SQ SEQUENCE 720 AA; 7F9B540A12A3DB2D CRC64;

Query Match 5.3%; Score 214.5; DB 2; Length 720;
Best Local Similarity 21.6%; Pred. No. 1.1e-06;
Matches 186; Conservative 101; Mismatches 315; Indels 259; Gaps 45;

QY 16 INTP-----LAQHETQSGVLEFVTVGVKSRPRATSGLLHTSTASDKIIS 62
Db 1 MNTPLFRLSLLSITLAAGFAHAENNAKVLDVTVTKGDRQ-----GSKIRNIV- 50
QY 63 GDTLRQK-----AVNLGDALGVGIHSAQVGGASAPVI-----RQGTERRIKV-----107

```

```

Db 51 --TLQQDBSTATDMRELLKEPSI---DFGGNGTTSQFLTRMGONGSNVDIKVDNAYS 105
QY 108 ---LNHGETGDMADFPDHAIMVDLTALSOQVBIIRGPVTLTYSSSGNAGLVADVADKIP 164
Db 106 SQILYHQGR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAIL 146
QY 165 EKMP-----NGVSGELGLRLSSG--NLEKLTSGGINIGLGNFVLHTEGLY-----RKSG 213
Db 147 AKTVDAQDLLKGLDKWGVRLNSGFSASNEGVSYGASVFGKEGNF-----DGLFSYNRNDEK 202
QY 214 DYAVPR-YRNL---KELPSP-----RRFANGQHRVILGWKRKYR-----251
Db 203 DYFAGKGFNRVNGKVPVPSALDKRSYLAKITGTFGDDHRIVLSHMKOQHRGIRTVREE 262
QY 252 -TVSDRRDQVGLPAHSHEDDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCG 310
Db 263 FTVGDKSSRINIDRQAPAYR-----TTQSNINLAYTGKNLG 299
QY 311 FHDDDDAHAAHANGKFWIDLNRKY---ELRAEWKQFPFGFEALRVHLNRNDVHHDEKAG 367
Db 300 FVEKLDANAYV-----LEKERYSADDSGTGYAGNVKGNPNHTRITTRGANFNFSRLA 351
QY 368 DAVENFNQOT-QNARIELRHO---PIGRKLGSGVQYLGQKSSALSATSEAVKQPM---420
Db 352 E-----QTLKYGINVRHQIKPOAFNLNSKESIPTTEEK-----CQKVDKPMEQQ 397
QY 421 -----LLDNKVQHYVSFFGVEQA--NWDNFTLEGGVVEKOKASIRYDKALIDR 466
Db 398 MKDRADEDTVHAYKLSNPTKTDITGVYVEAIHIDGFTLTGLAYDRFKVK--THDGKTVSS 456
QY 467 ENYKQPLDGLGAHQRTARSFALSQWNYFTPOHKLSTASHQERLSTQELYAHGKH---523
Db 457 SNLN-----PSFGVIWQPH-----HWSFSASHNY---ASRSPRL--YDALQTHGKGI 501
QY 524 -VATNTEFVGNKHLNKERSNNIELALGYEGDRWQYNLALYRNFNGVYIYAQTLNDGGRPK 582
Db 502 SIADGT-----KAERARNTIEGFNYNDGTFAAN-----GSYFW-QTIKDALANP 544
QY 583 SIEDDSEMKLVRYNQSADFYGAE-GEIYKPTPRYRIGVSGDYVRGLKNLPSLPGRD 641
Db 545 QNRHDSVAVREAVNAGYIKNHGYELGASVYRTGGLTAKVGVN-----HSKPRFYD 593
QY 642 AVGNRPFIADDDQNAAPRVPAARLGFHLKASLTDRI--DANLDY-YRVEAQNKLARYETRTP 699
Db 594 THKDKLLSANPEF-----GAQVGRWTASLAYRFQNPNLGWR-----GRYVQKAT 640
QY 700 GHMLNLGANYYR-----NTRYGEW-----NWYKADNLLNOSVYAHSS 738
Db 641 G-SILAAAGQKDRKGNLENVVRKGFVNDVFANWKPLGKDTLNVNLNVNVNFKFYPHSQ 699
QY 739 FLSDT-PQMGSRFTGGVNVKF 758
Db 700 RWTNTPGVRDVRLLGVNYKF 720

RESULT 31
Q9JWB8
ID Q9JWB8 PRELIMINARY; PRT; 714 AA.
AC Q9JWB8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative iron-regulated outer membrane protein.
GN FETA OR NMA0453.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,

```

RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagels K., Leather S., Moulé S., Mungall K., Quail M.A., Rajadream M.A., Ruthcroft K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.; "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.";
RT	Nature 404:502-506(2000).
RL	Nature 404:502-506(2000).
DR	EMBL; ALI62753; CAB83750.1; -.
DR	InterPro; IPR000531; TonB_boxC.
DR	Pfam; PF00593; TonB_dep_Rec.1.
DR	PROSITE; PS01156; TonB_DEPENDENT_REC_1.
KW	Complete proteome.
SQ	SEQUENCE 714 AA; 78885 MW; A3BBC1C1LED5527 CRC64;

Query Match	
Best Local Similarity 23.0%; Score 213.5; DB 16; Length 714;	
Matches 188; Conservative 103; Mismatches 308; Indels 257; Gaps 47	

Yy	11	LSILLINPL-LAQAHETEQSGLTETTVVKSRPRATSGLLHTSTASDKIIISGDTLRQK	69
Dd	8	LSLLSLTLAAGFAHAENANVALDITVKGDRQ-----GSKIRTNIV---TLQOK	55
Yy	70	---AVNLGDALDGVPGHASQYGGSAPVI-----RGTCRRTKV-----LNHH	111
Dd	56	DESTATDMRELLKEEPSI--DFGGNGTSGFTLVRGMGNSVDIKVNAYSDSOILYHQ	112
Yy	112	GETGMDAFSPDHAIWDITALSQVEIIRGPVTLLYSNGNAGLVADVADGKIPEKMPE--	169
Dd	113	GR-----FIVDPALKVVSVQKG-----AGSASAGIGATGAIAKTVDQAQ	153
Yy	170	---NGVSGELGRUSSG-NLEKLTSGGINTGLGNFVLHTEGLYRKS----GDYAVPR-	219
Dd	154	DLLXGLDKRWGVRLNSGFASFNEGVSYGASVFGEKNF---DGLFSYRNDEKOYEAKG	209
Yy	220	YRNL----KRLPDSP-----REFANGQHRAVLGWKKFRFYRYSDRDQY-GLP	263
Dd	210	FRNNFGKTPYPSALDKRSYLAKTGTFGGDDHRIVL-----SHMKOOHGRIR	258
Yy	264	AHSHEYDDCHADIITQKSLIN-KRYLIQLYPHLLITEEDVDYNPGLSCGFHDHDDAHAAH	322
Dd	259	TVREEF----TVGDKKERISMKQAQSYRE-TTOSNTLAYTKDLGFVEKLDNAVTV-	311
Yy	323	NGKPWIDLNRKY---ELRAEWKQFPFGFEALRVHLNRNDYHDEKAGDAVENFNENOT-	378
Dd	312	-----LEKKRYSADDSSGSGYAGNVKGFPHETQITTRGANFNFD SRLAE-----QTL	356
Yy	379	QNARIERHQ---PIGRLLKSGWGVOYLQCKSSAL-----SATSEAVKQPMLLDNKVQHY	429
Dd	357	LKYGINVRHQIKQAFINSQFKIE---DKGKATDEEKNKRENPKIAKAYRLTPTKTD	413
Yy	430	SFFGVEQAN-WDNFTLEGGVVEREKOKASIRYDKALIDRENYYKQPLPDGAHQRTARSFA	488
Dd	414	AGAYIEAHEIDGFTLTGLGYDRFKVKTDOG-----TVSSSS	452
Yy	489	LSGNW---YFTPOHKLSLTASHQ---ERLPSTOE-LYAHGKH---VATNTFVGNKHLNKE	539
Dd	453	LNPSFGVIWQPHEHWWSFSAASHNYASRSPRLDALQTHGKRGIIISTADGT-----KAE	504
Yy	540	RSNNIEALGEGDRWQYNLALYRNFONYLYAQTLNDG-RGPKSIEDDSEWK-----	591
Dd	505	RARTEIGFNTDGTFAAN-----GSYFW-QTIKDALNPQRHDSVAVREAVNAGY	555
Yy	592	-----LVRYNQSGADFAGEEIYFKPTPRYRIGVSGDYVRG---R	629
Dd	556	IKNHGYELGASYRTCGLTAKVGVSHSKPFYDTHDKLLSANPEGAQVGTWTASLAYR	615
Yy	630	LKNI.PSLP-----GR--EDAYGNRPFIADODQNAKPVPAARLGPHLKASLTDRIDANLDY	683
Dd	616	FKN-PNLEIGWRGVYQKAVSGSIIVAGKDRSGKLENVVRQGFGWN-----	660
Yy	684	RVFAQNKLARVETRTPGHHMLNLGANRPNTRYKGEWNVYKADNLLNOSVYAHSSF.SDT	743

```

Db      661 DVFANKPLGKDTLN-----VNLSVN-----NVNFTFFPHSQRWNT 699
Qy      744 -PQGRSFTGGVNVKF 758
Db      699 LPGVGDRVLRGVNVKF 714

RESULT 32
Q47232 PRELIMINARY; PRT; 673 AA.
AC Q47232;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Fyua precursor.
GN FYUA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli Phi;
RA Rakin A., Saken B., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence
RT factor with dual function.";
RL Mol. Microbiol. 13:0-0(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli Phi;
RX MEDLINE=95247659; PubMed=7730256;
DR EMBL; Z35106; CAA84489.1; -.
DR HSSP; P05825; 1PEP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep.Rec; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 673 FYUA.
SQ SEQUENCE 673 AA; 73665 MW; 218360F62C65C71F CRC64;

Query Match 5.3%; Score 213; DB 2; Length 673;
Best Local Similarity 21.8%; Pred. NO. 1.3e-06;
Matches 181; Conservative 93; Mismatches 323; Indels 232; Gaps 42

Qy 1 MAQTLTKPIVLSILLINTELLAQAHETQSQGLETVTVVQKSRPATSGLLHTSTASDKI 60
Db 1 MKMTRLPLALGGLL--PALNAQTSQD---ESTLVVTASKQSRSA--SANNVSTV 53
Qy 61 ISGDTLRQKAVNLGDALGV-PGIHASQYGGGASAPV-IRGTGRRIKVLNHHGTGDMA 118
Db 54 VSAPELSDAGVTASDKLPRLVPLGLNIENSGNWLFTSLRGVSSAQ----- 99
Qy 119 DF-SPDHAIWD-----TALS--QQVEILRGPVTLTYSSGNVAGLVADVADGKIPE 165
Db 100 DFYNPAVTLTYVDGVPQLSTNTIQALTVDQSVELLRGPQGTLYGRSAQGGINIVTQQ-PD 158
Qy 166 KMP----ENGVSQELGRLSSGNLEKLTSGGINIGLKNFVLHTSLGYRK--SGDYAVPR 219
Db 159 STPRGYIEGGVSSRDSYR-SKFNI-----SGFIQDGL-----LYGSTVLLRQVDDGDMINPA 209
Qy 220 YRN-----LKRLPD-----SPRFANGQRAVLGWRKGFYRFTYSDR 256
Db 210 TGSDDLGTTRASIGNVKLRPLADDPQWEMGFAASRECTRAQTQDAVGVN----- 258
Qy 257 RDQVGLPAHSHEYDDCHADIWQSLINKRVLQIYPHLLTREDVDYDNPGLSCGFHDDDD 316
Db 259 -----DINGKRLSISDGSPDPYMRCT-----DSQTLSGKYTTDD- 293

```

```
QY 317 AHAAHNGKPIDLNRKRYELAEWKO-----PFP-GFEALRVHLNRNDYHHDEKA---G 367
Db 294 -----WV-----FNLISAWQQQHYSTFPSSGLIVNIPORWNOVDQVELRAATLG 337
QY 368 DA--VENFNNTQARIELRHQPIGRKLGSGVQYVLGQKSSALSATSSEAVKQPMILLDNK 425
Db 338 DARTVDMVFLYRQNR-----EKLNSAYDMPYLSSTGYTTAETLAA----- 382
QY 426 VOHYSPFQVEQANW---DNFTLEGGVRVEKQKASIRYDKALIDRNNYKQPLDILGAHQK 482
Db 383 ---YS-----DLTWLTDTRFDITGGVRFVSHDKSSTQYHGSMLG-----NPFQDQKSD 428
QY 483 TARSFALSGNWYFTPHQKLSLTAHQERLPSQELYAHGKHVATNTFFVGNKHLNKRNS 542
Db 429 DQVLGQLSAGYMLTDDRWV-YTRVAQCYKPSGNI-----VPTAGLDAKFPVAKESI 479
QY 543 NIELALGYEGDRWQYNLALYRNRFGNYIYAQTLNDGRGPKSIEDDSEMKLVRYNOSGADF 602
Db 480 NYELGTRYE---TADVTLQAATF-YTHTKDQMLYSGPVGMQTLS-----NAGKADA 526
QY 603 YGAEGEIYFKPTPRYRIGYSGDYVRGRLLKNLPSLPQREDAYGNRPFIAQDDQNAFVPAA 662
Db 527 TGVLEAKRWFAPGWSWDINGNVIRSEFTN-----DSELYHGHR-----VPFVPRY 572
QY 663 RLGFHLKASLTDRIIDANLDYRVFAQNKLYARTETRPQGHMLNLGANYRNRTRYGE---- 718
Db 573 GAGSSVNGVIDTRYGA-----LMPRLAVNLVGFHYFD-GDNQLRQCTYATLDS 620
QY 719 --W-----NWVKADNLNQ--SVYAHSSFLSDTPQMGSRSTGGVNVK 757
Db 621 LGWQATERMNISVYVDNLFDRRYRTYGYVMNGSSAVAQVNMGRVTGINTR 669

RESULT 33
Q9JRC5 PRELIMINARY; PRT; 703 AA.
AC Q9JRC5;
DC 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Putative siderophore receptor (tonB-dependent receptor,
GN PHUA OR NMB0293.
OS Neisseria meningitidis, and
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487, 491;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20187481; PubMed=10722605;
RA Klee S.R., Nassif X., Kusecek B., Merker P., Beretti J.L., Achtman M.,
RA Tinsley C.R.;
RT "Molecular and biological analysis of eight genetic islands that
RT distinguish neisseria meningitidis from the closely related pathogen
RL neisseria gonorrhoeae."
RL Infect. Immun. 68:2082-2095 (2000).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=2017555; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science; 287:1809-1815 (2000).
```

```
DR EMBL; AJ391269; CAB72049.1; --
DR EMBL; AF002386; AAF40744.1; --
DR TIGR; NMB0293; --
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 703 AA; 78860 MW; 392CD05997A AF2FA CRC64;

Query Match 5.3%; Score 212.5; DB 16; Length 703;
Best Local Similarity 20.7%; Pred. No. 1.5e-06;
Matches 182; Conservative 117; Mismatches 260; Indels 321; Gaps 52;

QY 12 SILLINTPLLA---QAHETEQSVGLFTVTVCKSRPRATSGLLH-----TSTASDKLIISG- 63
Db 12 TLIIASFVPAADTQDNGEHETATLPTVSWVQSDTSVLKGVINTYDEAAVTRNGQLIKET 71
QY 64 ----DTLR-QKAVNLG-----DALDGVPGIHASQYGGGASAPVIRGOTGRRIKVLNHHGE 113
Db 72 PQTIDTINIKKNKNGTNDLSILLEGNAIDA-----AYDMRGES-----TFLRGF 117
QY 114 TGDMADFSPDHA-----IMVDTALSQQVEILRGVPTLLYSSGNVAGLVADVADGKIPEKMP 168
Db 118 QADASDIYRDGVRESGQVRSTANIERVEILKGPSSVLYGRNTGGGVINMVKYANFKQS 177
QY 169 EN--GVSG-----ELGLRLSS--GNLEKLTSG--GINIGLGNKPVLT 205
Db 178 RNIGAVYGSNWRSLNMDINEVLNKNVAIRLTGEVGRANSFRSGIDSKNMVSPSITVKL 237
QY 206 EGLYKSGDYAVPRYRNKRLPDP--SPRRFANQCHRAVLGWRKRFYRRTYSDDRDQVGLP- 263
Db 238 DNLGKWTGQYT---YDNVERTPDRSP-----TKSVY-----DRPLGY 272
QY 264 ----AHSHEYDDCHADIIWQKSLINKRYLQLYPHLLTBEEDVDYDNPGLSCGFHDDDDA 319
Db 273 RMGFAHRNDF-----VKDKLQWV-----RSDLEY----- 296
QY 320 HAHNGKPIDLNRKRYELAEWKOQFPFGFEALRVHLNRNDYH--DEKAGDAVENFP-- 374
Db 297 -AFNDK-W-----RAQWQ-----LAHRTAAQDFDHFYAGSENGLIKRNAYAM 336
QY 375 ---NNQTONVARIELR-HQPIGRKLG--SWGVOYLQK-----SSALSAT----- 412
Db 337 QQTDNKTLSSNLTLDNGDYTGIFENHLIVGMDYSREHRNPTLGFSSAFSASINPYDRASW 396
QY 413 -SEAVKQPMILLDN--KVQHYSPF--GVEQANW--NFTLEGGVRVEKQKASIRYDKALIDR 466
Db 397 PASGRLOPILITONRHKADSYGIFVQNIIFSATPDLKFLVG-----RYDKYTFNS 445
QY 467 ENYKQPLDILGAHQK--TARSFA--LSGNWYFTPHQKLSLTAHQERLPSQELYAHGKH 523
Db 446 ENKL-----TGSSRQYSGHSFSPNIGAVNINPVH--TLYASYNK--GFAPYGGRGY 494
QY 524 VATNTFEVGNKHLNKRNSNIELAL--GYEGDRWQYNLALYRNRFGNYIYAQTLNDGRGP 581
Db 495 LSIDTLSSAVFNADPEYTRQYETGVKSSWLDLRLSTLSAY----- 535
QY 582 KSIEDDSEMKLVRYNOSGADFYGAEGEIYFKPTPR---YRIGVSGD----- 624
Db 536 -----QIERFN-----IRYRDPKNNFYIYAVSGKHSRSGVLSAIGOI 574
QY 625 ----YVGRLLKNLPSLPQREDAYGNRPFIAQDDQNAFVPAAARLGFHLKASLTDRIIDAN 679
Db 575 IPKKLYLRGSLGWQAK-----VVEDKENPDRV-----GIHL--NNTSNVTGN 615
QY 680 LDYTRV-----FAQNKLYARTETRPQGHMLNLGANYRNRTRYGE 718
Db 616 L-FFRYTPETENLYGEIVGTGKRYGYNRSRKNKEVTTLGPFARVDAMLGNHKNVNTFA- 673
QY 719 WNWTVKADNLNQSVYAHSSFLSDTPQMGSRSTGGVNVK 758
Db 674 -----AANLILNQYRSDSM-----PGNPRGYTARVNYRF 703
```

```
RESULT 34
Q9JXL3 PRELIMINARY; PRT; 714 AA.
AC Q9JXL3;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Iron-regulated outer membrane protein FrpB.
GN NMB1988.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Ufferback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignan V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science. 287:1809-1815 (2000).
DR EMBL; AE002548; AAF42315.1; -.
DR TIGR; NMB1988; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Complete proteome.
SQ SEQUENCE 714 AA; 79150 MW; F8BAP5BDB14CDFE CRC64;

Query Match 5.3%; Score 212.5; DB 16; Length 714;
Best Local Similarity 21.8%; Pred. No. 1.5e-06;
Matches 187; Conservative 96; Mismatches 305; Indels 281; Gaps 47;

Qy 16 INTPL-----LAQAHETEQSGVLETVVVGKSRPRATSGLLHTSTASDKIIS 62
Db 1 MNTPLRLSLLSLTLAAGFAHAENNAKVLDTVTVGDRQ-----GSKRTNV- 50
Qy 63 GDTLRQK-----AVNLGALDGVPGIHASQYGGASAPVI-----RGOTGRRIKV----- 107
Db 51 --TLQKDESTATDNRELLKEPSI---DFGGNGTSQFLTRGMQNSVDIKVDNAYSD 105
Qy 108 ---LNNHGETGMADFSPDHAIMVDTALSOQVEILRGPVTLTLYSGNVAGLVVDVADGKIP 164
Db 106 SQILYHQGR-----FIVDPALVKVSVQKG-----AGSASAGIGATNGAI 146
Qy 165 EKMP-----NGVSGELGLRLSSG--NLEKLTSGGINIGLGNFVILHTEGLY-----RKSG 213
Db 147 TKTVDAQDLGLDKNGVRLNSGFASNGVSGASVFGKEGNF-----DGLFSYNNRNEK 202
Qy 214 DYAVPR-YRNL-----KRLPDSF-----RRFANGQHRVILGWKRFVRYRTYSDR 257
Db 203 DYACKGFRNPNFGKTVPYSDALKRSYLAIXIGTSFGDGRIVL-----SHMK 251
Qy 258 DQY-GLPAHSHEYDDCHADITWQSLIN-KRYLYLPHLLTEEDVDYDNPGLSGCGFHHDD 315
Db 252 DQHRGIRTVREF-----TVGGDKERISMERQAPAYRE-TTQSNNTLAYTCKNLGFVEKL 305
Qy 316 DAHAHANGKWDILRNKRYE-----LRAEWQPPFGFEAL 351
Db 306 DANAYV-----LEKRYSGADSGTGAGNVKPNHTQITTRGMNPNFDSRLAEQTL 357
Qy 352 RVHLNRNDYHDEKAGDAVENFFNNQTONARIELRHQPIGLKSGWGYQLGQSSALSA 411
Db 358 KYGIN---YRHOEIKPQA---FLNSQ---FKIEDKEKATDEKKN-----NR 395
Qy 412 TSEAVKQPMLLDNKVHYSFFGVQAN--WDNFTLEGGVVRVKQKASTRYDKALIDRENY 470
```

```
Db 396 ENEKIAKAYRLTNPTKTDGTGAYIEAIHEIDGFTLTGGLRYDRFKVK--THDGKTVSSNNLN 454
Qy 471 KOPLPDLAGHROTARSFALSQNNWTFPTQHKLSLTASHOERLPSTOELYAHGKH-----VAT 526
Db 455 ---PSFGVIWQPH-----HWSFSASHNY---ASRSPL--YDALQTHGKRGIIISAD 499
Qy 527 NTFEYVGNKHLNKRSSNNIELALGYEGDHWQYNLALYRNRFNGYIYAOTILNDG-RGPKSIE 585
Db 500 GT-----KAERARTEIGFNTYNDGTFAAN-----GSYFW-QTIKDALANPQNRH 542
Qy 586 DDEMK-----LVRYNOSGADFYGAEGEIVFKPFPYRI 619
Db 543 DSVAVREAVNAGYIKNHGYELGASVTCGLTAKGVSHSKPRFYDTHDKLLSANPEGA 602
Qy 620 GVSQDYVRG-----RLKNLPSP-----GR--EDAYGNRPPIAQQDQNAVPVPAARLGHK 670
Db 603 QVGRWTWASLAYRFQN--PNLEIGWRGRYVQKAVGSILVAGQKDRNGKLENVVRKGFQVNV- 660
Qy 671 SLTDRIANDLDYRVEAQNKLARYETRTPGHMLNGLNANYRNTRYGEMWVYKADNLLN 730
Db 661 -----DVFAWKPGLKDTLN-----VNLSVN-----NVFN 685
Qy 731 QSVYAHSGSFLSDT-PQMGSRFTGGVNVKF 758
Db 686 TFYYPHSQRWNTLPGVGRDVLGLVNYKF 714

RESULT 35
Q47231 PRELIMINARY; PRT; 673 AA.
AC Q47231;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Fyua precursor.
GN FYUA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli K49;
RX MEDLINE=95075311; PubMed=7984105;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence
RT factor with dual function."
RL Mol. Microbiol. 13:253-263 (1994).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli K49;
RX MEDLINE=95247659; PubMed=7730256;
RA Rakin A.V., Urbitsch P., Heesemann J.;
RT "Evidence for two evolutionary lineages of highly pathogenic Yersinia
RT species."
RL J. Bacteriol. 177:2292-2298 (1995).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Z38065; CAAB6212.1; -.
DR HSSP; P05825; 1FEP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 673 FYUA.
SQ SEQUENCE 673 AA; 73713 MW; 2635A53A7515BD08 CRC64;

Query Match 5.3%; Score 212; DB 2; Length 673;
Best Local Similarity 21.8%; Pred. No. 1.5e-06;
Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

Qy 1 MAQTTLPVLSILLINTPLIAQAHETEQSGVLETVVVGKSRPRATSGLLHTSTASDKI 60
```

```

Db 1 MKMTLYPLALGGLL--PTTANAQTSQD---ESTLVVTASKQSSRSA--SANNVSVTV 53
Qy 61 ISGDTLRQKAVNLGDALGV-PGIHASQYGGASAPV-IRGOTGRRIRKVLNHHGETGMA 118
Db 54 VSAPELSAGVTASDKLPRVLPLGLNIENSGNMLFSTISLRGVSSAQ-----99
Qy 119 DF-SPDHAIMVD-----TALS--QOVELIRGPVTLIYSSGNVAGLVADVADGKIPE 165
Db 100 DFYNPAVTLYVDGVPQLSTNTTQALTDVQSVELLRGPQGTLYGKSAQGGIINIVTQQ-PD 158
Qy 166 KMP-----ENGVSGLGLRLSSGNLEKLTSGGINIGLKNFVHLHTGLYRK--SGDYAVPR 219
Db 159 STPRGYIEGGVSSRDSYR-SKPNL-----SGPIQDGL-----LYGSVTLRQVDDGDMINPA 209
Qy 220 YRN-----LKRLPD-----SPRRFANGQHRVAVLGNRKRFRYRTYSR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPMWGFPAASRECTATQDAVGVN-----258
Qy 257 RDQYGLPAHSHEYDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDD 316
Db 259 -----DIIKGRKLSIDGSDPDYMRCT-----DSQTLGKVTDD-293
Qy 317 AHAHAHGKPMIDLNRKRYELRAEWKQ-----PFFGFEALRVHLNRNDYHHDEK-----365
Db 294 -----WV-----FNLSAQOQHYSTRFES-GSLIVNMPQR-WNQDVQELRAAT 335
Qy 366 AGDA--VENFFNNTQONARIELRHOPIGRKLSGWGVQYLGKSSALSATSEAVKQPMLLD 423
Db 336 LGDARTVDMVFLYRQNT--DIIKGRKLSIDGSDPDYMRCT-----DSQTLGKVTDD-293
Qy 424 NKVQHSYFFGVEQANW---DNFTLEGGVYRVEKQKASIRYDKALIDRENYKQPLDGLGAH 480
Db 393 -----YS-----DLTWHLTDTRFDIGGVRFHSHDKSSTQYHGSMLG-----NPFQDQGS 426
Qy 481 RQTARSAFALSGNWYPTPQHKLSLTASHQERLPSTQELYAHGKHVATNTPEVGNKHLNKR 540
Db 427 NDDQVLGQLSAGYMLTDDWRV-YTRVAQGYKPSGYNI-----VPTAGLDAKPFVAK 477
Qy 541 SNNIELALGYEGDRWQYNALYRNFNGVYIAQTLNDGGRPKSIEDDSEMKLVRYNQSA 600
Db 478 SINYELGTREY-----TADVTLOAATF--YTHTKDMQLYSGPGVMQTL--NAGKA 524
Qy 601 DFYGAEGEYIYKPTPRYRIGVSGDYVRGRKLNPLSLPGREDAYGNRPFFIAQDDQNAVRP 660
Db 525 DATGVELEAKWRFAPGWSWDINGNIRSEFTN-----DSELYHGNR-----VPFVP 570
Qy 661 AARLGFHLKASUTDRIDANLDYRVRFAQNKRLARYETRTFGHMLNIGANRYENTRYGE--718
Db 571 RYCAGSSVNGVIDTRYGA-----LMERLAVNLVGPYFD-GDNQLRQGTATLD 618
Qy 719 ---N---NWYKADNLNQ--SVVAHSSFLSDTPQMGSRFTGGVNVK 757
Db 619 SSLGWAQTERMNISVYVDMLFDRRYRTYGYMNGSSAVAQVNMGRVTGINTR 669

```

RESULT 36

```

Q57303
ID Q57303 PRELIMINARY; PRT; 673 AA.
AC Q57303;
DC 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE NCBI_TaxID=633, 562;
GN FYUA precursor.
GE FYUA
OS Yersinia pseudotuberculosis, and
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633, 562;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN-YERSINIA PSEUDOTUBERCULOSIS IA;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;

```

```

Mol. Microbiol. 13:0-0(1994).
[2]
SEQUENCE FROM N.A.
RC STRAIN-YERSINIA PSEUDOTUBERCULOSIS IA;
EX MEDLINE=95247659; PubMed=7730256;
RA Rakin A.V., Urbitsch P., Heesemann J.;
RT "Evidence for two evolutionary lineages of highly pathogenic Yersinia species.";
RL J. Bacteriol. 177:2292-2298(1995).
CC -|- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL: Z35107; CA884490.1; -
DR EMBL: Z35105; CA884488.1; -
DR HSP; F05825; IFRP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 673 FYUA.
SQ SEQUENCE 673 AA; 73682 MW; 5BFF31DA7EA15DFE CRC64;

Query Match
Best Local Similarity 5.3%; Score 212; DB 2; Length 673;
Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

Qy 1 MAQTTLKPIVILINTPLIAQAHEHETQSGVLEFVTVVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTLYPLALGGLL--PAIANAQTSQD---ESTLVVTASKQSSRSA--SANNVSVTV 53
Qy 61 ISGDTLRQKAVNLGDALGV-PGIHASQYGGASAPV-IRGOTGRRIRKVLNHHGETGMA 118
Db 54 VSAPELSAGVTASDKLPRVLPLGLNIENSGNMLFSTISLRGVSSAQ-----99
Qy 119 DF-SPDHAIMVD-----TALS--QOVELIRGPVTLIYSSGNVAGLVADVADGKIPE 165
Db 100 DFYNPAVTLYVDGVPQLSTNTTQALTDVQSVELLRGPQGTLYGKSAQGGIINIVTQQ-PD 158
Qy 166 KMP-----ENGVSGLGLRLSSGNLEKLTSGGINIGLKNFVHLHTGLYRK--SGDYAVPR 219
Db 159 STPRGYIEGGVSSRDSYR-SKPNL-----SGPIQDGL-----LYGSVTLRQVDDGDMINPA 209
Qy 220 YRN-----LKRLPD-----SPRRFANGQHRVAVLGNRKRFRYRTYSR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQPMWGFPAASRECTATQDAVGVN-----258
Qy 257 RDQYGLPAHSHEYDCHADIIWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFHDDDD 316
Db 259 -----DIIKGRKLSIDGSDPDYMRCT-----DSQTLGKVTDD-293
Qy 317 AHAHAHGKPMIDLNRKRYELRAEWKQ-----PFFGFEALRVHLNRNDYHHDEK-----365
Db 294 -----WV-----FNLSAQOQHYSTRFES-GSLIVNMPQR-WNQDVQELRAAT 335
Qy 366 AGDA--VENFFNNTQONARIELRHOPIGRKLSGWGVQYLGKSSALSATSEAVKQPMLLD 423
Db 336 LGDARTVDMVFLYRQNT--DIIKGRKLSIDGSDPDYMRCT-----DSQTLGKVTDD-293
Qy 424 NKVQHSYFFGVEQANW---DNFTLEGGVYRVEKQKASIRYDKALIDRENYKQPLDGLGAH 480
Db 383 -----YS-----DLTWHLTDTRFDIGGVRFHSHDKSSTQYHGSMLG-----NPFQDQGS 426
Qy 481 RQTARSAFALSGNWYPTPQHKLSLTASHQERLPSTQELYAHGKHVATNTPEVGNKHLNKR 540
Db 427 NDDQVLGQLSAGYMLTDDWRV-YTRVAQGYKPSGYNI-----VPTAGLDAKPFVAK 477
Qy 541 SNNIELALGYEGDRWQYNALYRNFNGVYIAQTLNDGGRPKSIEDDSEMKLVRYNQSA 600
Db 478 SINYELGTREY-----TADVTLOAATF--YTHTKDMQLYSGPGVMQTL--NAGKA 524
Qy 601 DFYGAEGEYIYKPTPRYRIGVSGDYVRGRKLNPLSLPGREDAYGNRPFFIAQDDQNAVRP 660
Db 525 DATGVELEAKWRFAPGWSWDINGNIRSEFTN-----DSELYHGNR-----VPFVP 570
Qy 661 AARLGFHLKASUTDRIDANLDYRVRFAQNKRLARYETRTFGHMLNIGANRYENTRYGE--718

```

```
Db 571 RYAGSSVNGVIDTRYGA-----LMRLAVNLVGHYFD-GDNQLRQGTATLD 618
Qy 719 ----W-----NWYKADNLLNQ--SVYAHSSFLSDTPQMGRSFTGGVNVK 757
Db 619 SSLGWQATERNISVYVDNLFDRRYRTYGYNGSSAVAQVNMGRVTGINTR 669

RESULT 37
Q47230
ID Q47230 PRELIMINARY; PRT; 673 AA.
AC Q47230;
DT 01-NOV-1996 (TremBrel. 01, Created)
DT 01-NOV-1996 (TremBrel. 01, Last sequence update)
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)
DE Fyua precursor.
GN Fyua.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli K235;
RX MEDLINE=95075311; PubMed=7984105;
RA Rakin A., Saken E., Harmsen D., Heesemann J.;
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence
factor with dual function.";
RL Mol. Microbiol. 13:253-263(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Escherichia coli K235;
RX MEDLINE=95247659; PubMed=7730256;
RA Rakin A.V., Urbitsch P., Heesemann J.;
RT "Evidence for two evolutionary lineages of highly pathogenic Yersinia
species.";
RL J. Bacteriol. 177:2292-2298(1995).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; Z38064; CA86211.1; -.
DR HSSP; P05825; IPEP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Reg; 1.
KW Membrane; Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 673 FYUA.
SQ SEQUENCE 673 AA; 73697 MW; C18FFB601EAL4110 CRC64;

Query Match
Best Local Similarity 22.1%; Pred. No. 1.5e-06; Length 673;
Matches 185; Conservative 96; Mismatches 306; Indels 252; Gaps 46;

Qy 1 MAQTLKPIVLSILLINTPLLAQAHEHQSVGLETVTVVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTRLYPALGGLL--PAIANAQTSQD---ESTLVVTASKQSSRSA--SANNVSTV 53
Qy 61 ISGDTLRQKAVNLGDGV--PGIHASQVGGASAPV--IRGOTGRIKVLNHHGTGMA 118
Db 54 VSAPELSAGVATSDKLPVLPVLPMIENSGNMLFSTISLRGVSSAQ----- 99
Qy 119 DF-SPDHAIWVD-----TALS--QQVEILRGPVTLTYSSGNVAGLVADVADGKIPE 165
Db 100 DFYNPAVTLVYDGVLPQSLTNTIQTALTDVQSVELLRGPQGLTKGSAQGGIINIVTQQ-PD 158
Qy 166 KMP-----ENGVSSELGLALSSGNLKLSTGGNIGLGRKFLVHTGLYRK--SGDYAVPR 219
Db 159 STPRGYIEGGVSSRDSYR-SKENL-----SGPIQDGL-----LYGSVTLRLRQVDGDMINPA 209
Qy 220 YEN-----LKRLED-----SPRRFANGQRAVLGWKRYRYRTYSR 256
Db 210 TGSDDLGTGRASIGNVKURLAPDDQFWEWGFPAASRECTRATQDAYVGNW----- 258
Qy 257 RDQYGLPAHSHEYDCHADIIMQKSLINKRYLQLYPHLLITBEDVDVDPNGLSCGFHDDDD 316
```

```
Db 259 -----DIKGRKLISDGSPPDPMRRC-----DSQTLSGKYTTDD- 293
Qy 317 AHAHAHNGKPMIDLRNKYELRAEWKQ-----PPFGFEALRVHLNRNDYHHDEK----- 365
Db 294 -----WV-----FNLSAQQQHYSKRTFPS--GSLIVNMPQR--WNQDVQELRAAT 335
Qy 366 AGDA--VENFFNQTQNIARIELRHQPIGRFLKSGWGVQYLQKSSALSATSSEAVKQPMLLD 423
Db 336 LGDARTVDMVFLYRQNTNR-----EKLSAYDMPMTPLYLSSTGYTTAETLAA----- 382
Qy 424 NKVQHSFFGVEQANW---DNFTLEGGVRVEKQASIRYDKALIDRENVYKQPLPOLGAH 480
Db 383 -----YS-----DLTWHLTDFDIGGVRFSHDKSKSTQYHGSMLG-----NPFQDQGS 426
Qy 481 RQTARSPALSGNWYFTPOHKLSTASHQERLPSTOELYAHGKHVATNTFEVGNKHLNKR 540
Db 427 NDDQVLGQLSAGYMLTDDWRV--YTRVAQGYKPSGYNI-----VPTAGLDAKPFVAK 477
Qy 541 SNNTELALGYEGDRWQYNLALYRNRFNGYIYAQTLDNGRGPKSIEDDSEMKLVRYNOSGA 600
Db 478 SINYELGTRYE---TADVTLOAATF--YTHKQMLQSYGPGVMQTLN-----NAGKA 524
Qy 601 DFYGAEGEYFEPKTPRYRIGVSGDYVRGRLKNLPSLPGREDAYGNR--PFIAQDDQNAAPRV 659
Db 525 DATGVLEAKWRFPAGWSWDINGNVIRSEFTN-----DSELVHGNNRVPV-----PRY 572
Qy 660 PAARLGFHLKASLDRIDANLDYRVFAONKLARYETRTP-----GHHMLNLGANYRR 712
Db 573 GAG-----SSVNGVID-----TRYGALMPRLALNLVGPVPHYFD-GDNQLR 610
Qy 713 NTRYCE-----W-----NWYKADNLLNQ--SVYAHSSFLSDTPQMGRSFTGGVNVK 757
Db 611 QGTATLDSLGWQATERMNISSVYVDNLFDRRYRTYGYNGSSAVAQVNMGRVTGINTR 669

RESULT 38
Q8CVY3
ID Q8CVY3 PRELIMINARY; PRT; 673 AA.
AC Q8CVY3;
DT 01-MAR-2003 (TremBrel. 23, Created)
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)
DE Putative pesticin receptor precursor.
GN C2436.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
KW Receptor; Complete proteome.
SQ SEQUENCE 673 AA; 73683 MW; 5BF31DA7EA15DFE CRC64;

Query Match
Best Local Similarity 21.8%; Pred. No. 1.5e-06; Length 673;
Matches 181; Conservative 96; Mismatches 318; Indels 236; Gaps 43;

Qy 1 MAQTLKPIVLSILLINTPLLAQAHEHQSVGLETVTVVVGKSRPRATSGLLHTSTASDKI 60
Db 1 MKMTRLYPALGGLL--PAIANAQTSQD---ESTLVVTASKQSSRSA--SANNVSTV 53
Qy 61 ISGDTLRQKAVNLGDGV--PGIHASQVGGASAPV--IRGOTGRIKVLNHHGTGMA 118
```

Db 54 VSAPELSDAGVTASDKLPRVLPGIENSGNMLFSTISLGVSSAQ----- 99
QY 119 DP-SPHAIMVD-----TALS-QQVEILRGVPTLLSYSSGNVAGVVDVADGKPE 165
Db 100 DFYNPAVTVYDGVPLQSTNTIQALTDVSGVELLRPGQGLTYGKSAQGGIINVTQO-PD 158
QY 166 KWP-----ENGVSGLGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRK--SGDYAVPR 219
Db 159 STPRGVEGVSRSYSR-SKFLN-----SGPIQDGL-----LYGSVILLKQVDDGDMINPA 209
QY 220 YRN-----LKELPD-----SPRRFANGQRAVLGRRKRFYRRYTSR 256
Db 210 TGSDDLGGTRASIGNVKRLAPDDQWEMGFASRECTRATQAYVGN----- 258
QY 257 RDQYGLPAHSHEYDDCHADIIWOKSLINKRYLOLYPHLLTEEDVDVNDPGLSCFFHDDDD 316
Db 259 -----DIKGRKLSIDGSPDPYMRCT-----DSQTLISKYTTDD-- 293
QY 317 AHAHANGKFWIDLRNKRVELRAEWKQ-----PFGFPEARLVRHLNRNDVHHDEK----- 365
Db 294 -----WV-----FNLISAWQQOHSYRTPFS--GSLIVNMPOR--WQDVQOELRAAT 335
QY 366 AGDA--VENFFNNQTNARIELRHQPIGRKGSWGVOYLQKSSALSATSEAVKQPMLLD 423
Db 336 LGDARTVDMVFGLYRQNT-----EKLNSAYDMPYLSSTGYTTAETLAA----- 382
QY 424 NKVQHVSFPGVEQANW---DNFTLEGGVVEKOKASIRYDKALIDRENYKQPLDGLAH 480
Db 383 -----YS-----DLTWHLTDRFDIGGVRFSHPDKSSQVHSGMLG-----NPFQDGKS 426
QY 481 ROTARFALSNGWYFTPOHKLSLTASHQBELPSTQELYAHGKHVAATNTEVGNKHLNKR 540
Db 427 NDDQVLGQLSAGYMLTDDMRV-YTRVAQGYKPSGYNI-----VPTAGLDKAPFAEK 477
QY 541 SNNIELALGYEGRWQYNLALYRNGYIYAOTLNDGRGPKIEODSEMKLVRYNQSGA 600
Db 478 SINYELGTRYE---TADVTLQAATF--YTHTKDMQLYSGPVGQMOTLS-----NAGKA 524
QY 601 DFYGAEGEYFKPTPRYRGSDYVGRKLNPLSPGREDAYGNRPFFIAQDDQNAPRVP 660
Db 525 DATGVELEAKRFPAGWSWDINGNVRISFTN-----DSELYHGNR-----VFFVP 570
QY 661 AARLGFHLKASLTDRIDANDLYYRFAQNKRLARYETRTGHHMLNLGANYRRNTRYGE-- 718
Db 571 RYGAGSSVNGVIDTRYGA-----LMPLRLVNLVGPHYFD-GDQLRGQTYATLD 618
QY 719 -----W-----NWYVADMLLNO--SVYAHSSFLSDTPQGRSFTGQNVK 757
Db 619 SSLGWQATERMNISSVYVDNLFDYRRYTYGYMNGSSAVAQVMGRITVGINTR 669

RESULT 39

Q9XD18 PRELIMINARY; PRT; 687 AA.
AC Q9XD18;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative TonB-dependent outer membrane receptor protein.
OS Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=NTC 9343;
RX MEDLINE=99307214; PubMed=10377135;
RA Comstock L.B., Coyne M.J., Tzianabos A.O., Pantosti A.,
RA Oerendonk A.B., Kasper D.L.;
RT "Analysis of a capsular polysaccharide biosynthesis locus of
RT Bacteroides fragilis";
RL Infect. Immun. 67:3525-3532 (1999).
-|- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).

DR EMBL; AF048749; AAD40727.1; --
DR InterPro; IPR005531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KW Membrane; Outer membrane; Receptor; TonB_box.
SQ SEQUENCE 687 AA; 78838 MW; BBEB5E3502815E59 CRC64;
Query Match 5.2%; Score 211.5; DB 2; Length 687;
Best Local Similarity 22.1%; Pred. NO. 1.7e-06;
Matches 142; Conservative 84; Mismatches 220; Indels 197; Gaps 32;
QY 9 IVLSILLINT-PLLAQAHETEQSVGLETVTVGKSPRATSGLLTSTASDK----- 59
Db 6 LALLTLIGTLPLAAQNVREQ-----DVSVMND--PFNLEQIVVTATREKIKNTPVI 60
QY 60 --IISGDTLRQXAV-NLGDAL-DGVFGIHAHQYGGGASAPVIRGQTRRIKVL-----NHH 111
Db 61 TQIITSKQIEERGNTGNIQDLLTQEVPLNFQEVGYGTSID-IQGLGSKHILFLIDGERIA 119
QY 112 GETGMDAESPDAIMVDTALSOQVEILRGPTVLLSYSSGNVAGLVADVADGKIPEKM----- 167
Db 120 GENGGNIDYSRLNLYNID-----HIEIVKGASSALYGSQAMGGVINIITRKAKKFEASA 174
QY 168 -----PENGVSGBELGRLSSGNLEKLTSGGINIGLGNFVLHTEGLYRKS 212
Db 175 GIRYAGNQYKDTFKDHSQYKRIHLDPNLTNLSGLNLG---KFTMTDVLKSF 231
QY 213 GYAV-----PRY-----RNKRLPDSRRFANGQRAVLGWRKRFYRTYSDRR 257
Db 232 DGQLPDKKPLVKYFPAYNTTITELSKTPTSISGYEDVQVAHKMDY--RFSKRLKVQLK 289
QY 258 DOYGLPAHSHEYDDCHADIIWOKSLINKRYLOLYPHLLTEED-----VDY---DNPGLS 308
Db 290 GSYM--LNVY--DFQADNIFEKS-----EDTYGGSIDYTTISDKSSLV 329
QY 309 CGFHDDDDAHAHANGKFWIDLRNKRVELRAEWKQFPFGPEARLVRHLNRNDVHHDEKAGD 368
Db 330 ASVHTD-----HYNRYD-KVELKSGR 349
QY 369 AVENFFNNQTNARIELRHQPIGRKGSWGVOYLQKSSALSATSE-AVKQPMLLDNKVQ 427
Db 350 RLE--YKNNIIOPRIYVSTTALDKQITIGGLEI--YRESLFSDFKFTGVK-----ENKSQ 400
QY 428 HYSFFGVEQANWD---NFTLEGGRVE-----KQKASIRYD-KALIDRENYKQPL 474
Db 401 WTA-TAFLQDDWSINKQFSVIAGLRCDYHEKYGTNLTTPKASVMYKIFPFTVRFNY----- 454
QY 475 POLGAHQRTARSPALSGNWYFTPOHKLSLTASHQBELPSTQELYAHGKHVAATNTEVGNK 534
Db 455 -----ARGY-----RSPSIKELYMNWDHLGM-FWIYVGS 482
QY 535 HLKERSNNIELALGYEGRWQYNLALYRNGYIYAOTLND 577
Db 483 KLPETNNYISLGSYVNSWINANVNSWFRNKIEGMSND 525
RESULT 40
Q51132 PRELIMINARY; PRT; 697 AA.
ID Q51132
AC Q51132;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Iron-regulated outer membrane protein FtpB (fragment).
GN FtpB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2996;
RX MEDLINE=97124210; PubMed=8969523;
RA Der Ley P., der Biesen R., Suttmuller R., Hoogerhout P., Poolman J.T.;

RL
Microbiology 142:3269-3274 (1996)

Matches 182; Conservative 88; Mismatches 314; Indels 261; Gaps 42;

26 ETEOSVGL.EVTWVCKSRPRATSGLI.IHTSTASDKTISGDTT.BOK---AVNT.GNAT.NGVP 81

QV 82 GIHASOYGGGASAPVI-----RGOTGRRIKV-----LNHHGETGMDAFSPDHA1M 12

Qy 128 VDTALSOOVEILRGPTLLYSSGNVAGLVDADGKIPEKMP-----NGVSGELGLRLSS 182

Qy 183 G--NLEKLTSGGINIGLGKNFVLHTEGLY----RKSGDYAVPR-YRNL---KRLPDSP-- 230

Qy 231 - - - - -RRFANGQHRAVLGWRKRFYRTYSDDRRDQY-GLPAHSHEYDDCHADI IWOK 280

Qy 281 SLINKRYLQLYPHL-----LTEEDVDYDNPLSCGFHDDDDAHAAHNGKPWIDLNRKRY 339

QY 336 ELRAE--WKQPPGFALRVHLNRNDYHHDEKAGDAVENFFNNQT-QNARIELRHO---388

QY 389 PIGRLKGSWGVQYLGQKS-----SALSATSEAVKQPMLLDNKVQHYSHFFGVEOAN 438

QY 439 WDN-FTLEGGVRVEKQKASIRYDKALIDRENYKQPLDLAGHRQTARSFALSGNWFYTP 49

QY 498 QHKLSTASHQERLPSTQELYAHGKHVATNTFEVGNKHLNERSNNIELALGYEGDRWQY 55

QY 558 N-----LALYRNRFNGNYIYAQT LNDG-----RGP 58J

UY
582 KSIEDDSEMKLVRYNQSGADFYGAE-----GEIYFKPTPRYRIGVSGDYVRGRLKNLP 634

QY 635 SLPGREDAIGNRFFLAQUDUNAPRVPAARLGFHLKASLTDRI DANLDYRVFAQNKLARY 694

093 EIKIFGHHNLTGANYIKRNIRIGENWYVKADNLTNLSVIAHSSFLSDT-PQMGRSFTGG 753

QY
/34 VNVNF /38
|||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:00:43 ; Search time 47 Seconds
(without alignments)
2559.888 Million cell upd

Title: US-09-936-377-2

Perfect score: 758

Sequence: 1 MAQTTLKPIVLSILLINTPL.....FLSDTPQMGRSFTGGVNVKF 758

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters:	1107863
--	---------

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : A Geneseq 19Jun03:*

1:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSL1/cgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %			ID	Description
		Match	Length	DB		
1	758	100.0	758	21	AAB18719	A. Neisseria mening
2	286	37.7	339	21	AAV74562	Neisseria meningit
3	286	37.7	735	21	AAV74556	Neisseria meningit
4	286	37.7	758	23	AAU73011	Neisseria meningit
5	192	25.3	758	21	AAV74565	Neisseria meningit
6	175	23.1	339	21	AAV74563	Neisseria meningit
7	175	23.1	764	21	AAV74557	Neisseria meningit
8	175	23.1	764	23	ABB78068	Amino acid sequenc
9	138	18.2	728	24	ABP80923	N. gonorrhoeae ami

521	6	0.8	69	22	ABG23282	Novel human diagno	594	6	0.8	88	22	AAO13725	Human polypeptide
522	6	0.8	70	22	AAU62960	Propionibacterium	595	6	0.8	88	23	ABP04898	Human ORFX protei
523	6	0.8	70	23	ABP35014	Human ORF3987 prot	596	6	0.8	89	21	ABG40445	Arabidopsis thalia
524	6	0.8	71	21	AAU53851	Human colon cancer	597	6	0.8	89	22	ABG18696	Novel human diagno
525	6	0.8	71	22	AAU54956	Propionibacterium	598	6	0.8	90	22	ABG27030	Novel human diagno
526	6	0.8	71	22	AAU63746	Propionibacterium	599	6	0.8	90	22	AAU22539	Novel human colon
527	6	0.8	71	23	ABP06343	Human ORFX protein	600	6	0.8	90	22	AAU92496	Human digestive sy
528	6	0.8	72	20	AAU59922	Human myometrium t	601	6	0.8	91	21	AAU57594	Arabidopsis thalia
529	6	0.8	72	21	AAU07879	Arabidopsis thalia	602	6	0.8	91	21	AAU57993	Arabidopsis thalia
530	6	0.8	72	21	AAU32901	Arabidopsis thalia	603	6	0.8	91	21	AAU57993	Arabidopsis thalia
531	6	0.8	72	21	AAU45329	Arabidopsis thalia	604	6	0.8	91	21	AAU61460	Arabidopsis thalia
532	6	0.8	72	22	AAU44084	Propionibacterium	605	6	0.8	91	22	ABG07512	Novel human diagno
533	6	0.8	72	22	ABG11779	Novel human diagno	606	6	0.8	91	22	AAU29630	Novel human secret
534	6	0.8	72	23	ABP07293	Human ORFX protein	607	6	0.8	91	22	AAU09573	Human polypeptide
535	6	0.8	72	23	ABU53109	Human ORF15 protei	608	6	0.8	91	23	ABG64776	Human albumin fusi
536	6	0.8	73	15	AAU49660	Sequence of Heliot	609	6	0.8	91	23	AAU17185	Human secreted pro
537	6	0.8	73	20	AAU23076	T. gondii immunoge	610	6	0.8	91	23	AAU17215	Human secreted pro
538	6	0.8	73	22	AAU42135	Propionibacterium	611	6	0.8	92	20	AAU11433	Human 5' ESR secre
539	6	0.8	73	22	ABG04681	Novel human diagno	612	6	0.8	92	21	AAU40444	Arabidopsis thalia
540	6	0.8	73	22	ABU03925	Human musculoskele	613	6	0.8	92	22	ABG15046	Novel human diagno
541	6	0.8	73	22	AAU25547	T. gondii immunoge	614	6	0.8	92	22	ABG16016	Novel human diagno
542	6	0.8	73	24	ABU13219	Novel human muscul	615	6	0.8	92	22	ABG27579	Novel human diagno
543	6	0.8	74	22	AAU55320	Propionibacterium	616	6	0.8	92	23	ABP55320	Human cell signall
544	6	0.8	74	22	ABU10404	Human cDNA SEQ ID	617	6	0.8	92	23	ABP03378	Human ORFX protei
545	6	0.8	74	22	AAU18095	Novel human uterin	618	6	0.8	93	23	ABF42406	Human ovarian anti
546	6	0.8	74	22	AAU15992	Human novel secret	619	6	0.8	94	21	AAU17793	Arabidopsis thalia
547	6	0.8	74	22	AAU19911	Novel human calci	620	6	0.8	94	21	AAU57047	Arabidopsis thalia
548	6	0.8	74	23	ABP66991	Human polypeptide	621	6	0.8	94	21	AAU61450	Arabidopsis thalia
549	6	0.8	74	23	ABU05722	Novel human protei	622	6	0.8	94	22	AAU42596	Propionibacterium
550	6	0.8	74	23	ABP40775	Staphylococcus epi	623	6	0.8	94	22	ABU16469	Human nervous syst
551	6	0.8	75	22	AAU47596	Propionibacterium	624	6	0.8	94	22	AAU02044	Human polypeptide
552	6	0.8	75	22	AAU51458	Propionibacterium	625	6	0.8	94	22	AAU08326	Human polypeptide
553	6	0.8	75	22	AAU52259	Propionibacterium	626	6	0.8	94	23	ABP04847	Human ORFX protei
554	6	0.8	75	22	AAU5382	Novel human diagno	627	6	0.8	95	22	AAU58826	Propionibacterium
555	6	0.8	76										

813	6	0.8	123	21	AAG44953	Zea mays protein f	886	139	22	ABG03688	Novel human diagno
814	6	0.8	123	22	AAO06597	Human polypeptide	887	139	22	AAE03783	Human gene 1 encod
815	6	0.8	123	22	ABU05142	Pig galanin relate	888	139	22	AAE01703	Human gene 4 encod
816	6	0.8	123	23	ABU10562	Galanin-like pepci	889	139	23	ABG63979	Human albumin fusi
817	6	0.8	124	21	AAG13062	Arabidopsis thalia	890	139	23	ABG65498	Human albumin fusi
818	6	0.8	124	21	AAG16950	Arabidopsis thalia	891	140	22	ABU71829	Infectious salmon
819	6	0.8	124	21	AAG13859	Arabidopsis thalia	892	140	24	ABB84620	Human wild-type VE
820	6	0.8	124	23	ABP00672	Human ORFX protein	893	140	24	ABG73749	Human VEGF-C monom
821	6	0.8	125	13	ABP24512	Prod. of UCD-2 of	894	141	20	AAU37022	Protein which is s
822	6	0.8	125	14	ABU24512	Feline infectious	895	141	21	AAU44952	Zea mays protein f
823	6	0.8	125	14	ABU24512	Arabidopsis thalia	896	142	20	AAU85738	Polypeptide with t
824	6	0.8	125	22	AAU19592	Propionibacterium	897	142	20	AAU89900	Antigen 2 from clu
825	6	0.8	126	21	AAU32296	Arabidopsis thalia	898	142	21	AAU57323	Arabidopsis thalia
826	6	0.8	126	21	AAU32296	Arabidopsis thalia	899	142	22	AAU25727	Breast cancer-asso
827	6	0.8	126	21	AAU32296	Arabidopsis thalia	900	142	22	AAU05584	Human polypeptide
828	6	0.8	126	21	AAU32296	Polypeptide encode	901	142	23	ABU89954	Human breast cance
829	6	0.8	126	22	AAU08409	Polypeptide encode	902	142	23	AAU35954	Human polypeptide
830	6	0.8	126	22	AAU08410	Polypeptide encode	903	144	22	ABU95880	Extended human sec
831	6	0.8	126	22	AAU08412	Polypeptide encode	904	144	22	AAU43167	Human testicular a
832	6	0.8	126	22	AAU08413	Polypeptide encode	905	144	22	AAU95180	Propionibacterium
833	6	0.8	126	22	AAU08414	Polypeptide encode	906	144	22	AAU95180	Human reproductive
834	6	0.8	127	21	AAU41278	Human ORFX ORF1042	907	145	21	AAU43092	Human ORFX ORF2856
835	6	0.8	127	22	ABU08405	Polypeptide encode	908	145	22	AAU06479	Human polypeptide
836	6	0.8	127	23	ABU03223	Human ORFX protein	909	145	23	ABU0787	Magnetic particle
837	6	0.8	127	23	ABU10027	Insulin-like growt	910	146	21	AAU16299	Eucalyptus grandis
838	6	0.8	128	21	AAU58628	Arabidopsis thalia	911	146	21	AAU16299	Arabidopsis thalia
839	6	0.8	128	22	AAU08415	Polypeptide encode	912	146	21	AAU16299	Arabidopsis thalia
840	6	0.8	128	22	AAU08416	Polypeptide encode	913	146	22	AAU93192	Human protein sequ
841	6	0.8	128	22	AAU08417	Polypeptide encode	914	146	22	AAU94383	Human protein sequ
842	6	0.8	128	22	AAU08418	Polypeptide encode	915	146	22	AAU94403	Human protein sequ
843	6	0.8	128	22	AAU08419	Polypeptide encode	916	146	22	AAU94983	Human protein sequ
844	6	0.8	128	22	AAU08421	Polypeptide encode	917	146	22	AAU48932	MuCA protein. Uni
845	6	0.8	129	21	AAU66599	Arabidopsis thalia	918	147	22	ABU71594	Drosophila melanog
846	6	0.8	129	21	AAU66599	Human secreted pro	919	147	23	ABU70151	Human prey protein
847	6	0.8	129	22	ABU07629	Novel human diagno	920	147	23	ABU89291	Human polypeptide
848	6	0.8	129	22	AAU08406	Polypeptide encode	921	148	21	AAU09454	Arabidopsis thalia
849	6	0.8	130	20	AAU01485	ApM1 protein fragm	922	148	22	AAU52382	Arabidopsis thalia
850	6	0.8	130	20	AAU01486	acrip30 protein fra	923	148	22	AAU021241	Novel interleukin-
851	6	0.8	130	20	AAU01487	Arabidopsis thalia	924	149	21	AAU40220	Human secreted pro
852	6	0.8	130	21	AAU04561	Arabidopsis thalia	925	149	24	ABU75640	Human secretory po
853	6	0.8	131	21	AAU57324	Arabidopsis thalia	926	150	19	AAU83187	Staphylococcus aur
854	6	0.8	131	22	AAU23233	Novel human enzyme	927	151	21	AAU16268	Pinus radiata AGP
855	6	0.8	131	22	ABU51489	Helicobacter pylor	928	151	21	AAU18952	Zea mays protein f
856	6	0.8	131	23	ABU08216	Human ORFX protein	929	151	21	AAU44246	Arabidopsis thalia
857	6	0.8	132	17	ABU92117	Human Pac-1. Homo	930	151	21	AAU59845	Arabidopsis thalia
858	6	0.8	132	21	AAU07123	Arabidopsis thalia	931	151	21	AAU60285	Arabidopsis thalia
859	6	0.8	132	22	AAU078432	Human protein SEQ	932	151	22	ABU16240	Novel human diagno
860	6	0.8	132	22	AAU04159	Human polypeptide	933	151	22	ABU12182	Human PRO4990 poly
861	6	0.8	133	21	AAU86482	Human gene 54-enco	934	151	24	ABU66580	Human secreted/tra
862	6	0.8	133	22	AAU32290	Novel human secret	935	151	24	ABU66856	Human secreted and
863	6	0.8	133	22	AAU010761	Human polypeptide	936	151	24	ABU59661	Novel secreted aur
864	6	0.8	134	20	AAU36108	Extended human sec	937	152	19	AAU77504	Human endometrium
865	6	0.8	134	21	AAU13061	Arabidopsis thalia	938	152	20	AAU60208	Extended human sec
866	6	0.8	134	21	AAU15718	Arabidopsis thalia	939	152	20	AAU53963	Human SEEK1 protei
867	6	0.8	134	21	AAU21858	Arabidopsis thalia	940	152	23	ABU53108	Human SEEK1 wild t
868	6	0.8	134	21	AAU44247	Arabidopsis thalia	941	152	23	ABU53109	Human SEEK-1 prote
869	6	0.8	134	21	AAU54793	Arabidopsis thalia	942	152	23	AAU79581	Human SEEK-1 prote
870	6	0.8	135	22	AAU62140	Propionibacterium	943	152	23	AAU79760	Bcl-XL. Homo sapi
871	6	0.8	135	22	ABU08550	Novel human diagno	944	152	24	ABU38913	Staphylococcus epi
872	6	0.8	135	22	AAU08879	Human polypeptide	945	154	21	ABU10792	M. tuberculosis lu
873	6	0.8	135	22	AAU020661	Human secreted pro	946	154	22	AAU50374	Propionibacterium
874	6	0.8	136	21	AAU07644	Arabidopsis thalia	947	154	22	AAU50374	Human protein sequ
875	6	0.8	136	21	AAU49504	Arabidopsis thalia	948	154	22	AAU50374	Amino acid sequenc
876	6	0.8	136	21	AAU77977	A. thaliana enviro	949	155	18	AAU27977	LOX-1 isoform 9.
877	6	0.8	136	22	ABU16919	Novel human diagno	950	155	24	AAU79880	Arabidopsis thalia
878	6	0.8	137	17	ABU86948	PR35 traA fusion	951	156	21	AAU11340	Novel human diagno
879	6	0.8	137	22	ABU02021	Novel human diagno	952	156	22	AAU24219	Human immune/haema
880	6	0.8	137	22	ABU24915	Novel human diagno	953	157	20	AAU88001	Protein encoded by
881	6	0.8	137	22	AAU81505	S. epidermidis ope	954	157	22	AAU88001	C glutamic prote
882	6	0.8	137	24	ABU76268	Human GENSEP prote	955	157	22	AAU89846	Human BAFF recepto
883	6	0.8	137	24	ABU99359	Amino acid sequenc	956	157	22	AAU60700	Corynebacterium gl
884	6	0.8	138	17	AAU92524	Pyrodictium occult	957	158	17	AAU94674	Flea saliva protei
885	6	0.8	138	21	AAU36064	Zea mays protein f	958	158	23	ABU05420	M. tuberculosis an

QY 541 SNNIELALGVGDRWOYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVYNSGSA 600
 Db 541 SNNIELALGVGDRWOYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKLVYNSGSA 600
 QY 601 DFYGAEGEYFKEPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFFIAQDDQONAPRVP 660
 Db 601 DFYGAEGEYFKEPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFFIAQDDQONAPRVP 660
 QY 661 AARLGFHLKASLTDRIANLDYRVFAQNKLYARETRTPGHMNLNGANYRRNTRYGEWN 720
 Db 661 AARLGFHLKASLTDRIANLDYRVFAQNKLYARETRTPGHMNLNGANYRRNTRYGEWN 720
 QY 721 WYVKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 758
 Db 721 WYVKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 758

RESULT 2
 AAY74562
 ID AAY74562 standard; Protein; 339 AA.
 XX
 AC AAY74562;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:598.
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR N-PSDB; AA253324.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 419- ; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. the polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 339 AA;
 Query Match 37.7%; Score 286; DB 21; Length 339;
 Best Local Similarity 100.0%; Pred. No. 3.8e-291;
 Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 473 PLPDLGAHQRTARSAFALSGNWYFTPOHKLSTASHOERLPSTQELYAHGKHVATNTPEVG 532
 Db 54 PLPDLGAHQRTARSAFALSGNWYFTPOHKLSTASHOERLPSTQELYAHGKHVATNTPEVG 113
 QY 533 NKHLNKRSSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKL 592
 Db 114 NKHLNKRSSNNIELALGYEGDRWQYNLALYRNFGNYIYAQTLDNGRGPKEIEDDSEMKL 173
 QY 593 VRYNQSADPYGAEGEYFKEPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFFIAQD 652
 Db 174 VRYNQSADPYGAEGEYFKEPTPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPFFIAQD 233
 QY 653 DQNAFVPAARLGFHLKASLTDRIANLDYRVFAQNKLYARETRTPGHMNLNGANYRR 712
 Db 234 DQNAFVPAARLGFHLKASLTDRIANLDYRVFAQNKLYARETRTPGHMNLNGANYRR 293
 QY 713 NTRYGWNWYVKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 758
 Db 294 NTRYGWNWYVKADNLLNQSYAHSSFLSDTPQMGSRFTGGVNVKF 339

RESULT 3
 AAY74556
 ID AAY74556 standard; Protein; 735 AA.
 XX
 AC AAY74556;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 147 protein sequence SEQ ID NO:586.
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR N-PSDB; AA253318.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

XX Claim 2; Page 413; 1453pp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AA254616 to AA254617 represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides and polypeptides. AA254537 to AA254576 and AA254616 to AA254617 represent PCR primers used in the exemplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to *Neisseria meningitidis* (e.g. meningitis and septicaemia), to detect the presence of *Neisseria meningitidis*, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.

XX SQ Sequence 735 AA;

Query Match 37.7%; Score 286; DB 21; Length 735;
Best Local Similarity 100.0%; Pred. No. 8e-291;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 PLPLGHAHROTARSAFALSGNWFYTPQHKLSTASHQERLPSTQELYAHGKHVAINTFEVG 532
Db 450 PLPLGHAHROTARSAFALSGNWFYTPQHKLSTASHQERLPSTQELYAHGKHVAINTFEVG 509
Qy 533 NKHLNKRSNNIELALGYEGDRWQYNLALYRNFRGNFYIAQTLDNGRGPKEIEDDSEMKL 592
Db 510 NKHLNKRSNNIELALGYEGDRWQYNLALYRNFRGNFYIAQTLDNGRGPKEIEDDSEMKL 569
Qy 593 VRYNQSADFYGAGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 652
Db 570 VRYNQSADFYGAGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 629
Qy 653 DONAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRR 712
Db 630 DONAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRR 689
Qy 713 NTRYGEWNWYVKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 690 NTRYGEWNWYVKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 735

RESULT 4
AAU73011 standard; Protein; 758 AA.
AAU73011;
12-MAR-2002 (first entry)
Neisseria meningitidis virulence protein #101.
Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
infection; Gram-negative bacteria; antimicrobial.
Neisseria meningitidis.
WO200185772-A2.
15-NOV-2001.
08-MAY-2001; 2001WO-GB02003.
08-MAY-2000; 2000GB-0011108.
(MICR-) MICROSCIENCE LTD.
Tang C;
WPI; 2002-066593/09.

DR N-PSDB; AAS97296.
XX New peptide encoded by operon including virulence genes of *Neisseria meningitidis*, useful as vaccine component for treating or preventing meningitis and for identifying antimicrobial drug -
PS Claim 4; Page 412-415; 423pp; English.
XX The invention relates to a peptide (I) encoded by an operon (II) of *Neisseria meningitidis* including virulence genes, or a related molecule having a 40% sequence similarity at the peptide or nucleotide level in a Gram-negative bacterium, or its functional fragment, for therapeutic or diagnostic use. (I) and (II) are useful in the manufacture of a medicament for treating or preventing a condition (e.g., meningitis) associated with infection by *Neisseria* or Gram-negative bacteria. The product is useful for veterinary treatment and in a screening assay for the identification of an antimicrobial drug. The vaccines have prophylactic applications. AAU72911-AAU73014 represent *N. meningitidis* virulence proteins of the invention.

XX SQ Sequence 758 AA;

Query Match 37.7%; Score 286; DB 23; Length 758;
Best Local Similarity 100.0%; Pred. No. 8.2e-291;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 PLPLGHAHROTARSAFALSGNWFYTPQHKLSTASHQERLPSTQELYAHGKHVAINTFEVG 532
Db 473 PLPLGHAHROTARSAFALSGNWFYTPQHKLSTASHQERLPSTQELYAHGKHVAINTFEVG 532
Qy 533 NKHLNKRSNNIELALGYEGDRWQYNLALYRNFRGNFYIAQTLDNGRGPKEIEDDSEMKL 592
Db 533 NKHLNKRSNNIELALGYEGDRWQYNLALYRNFRGNFYIAQTLDNGRGPKEIEDDSEMKL 592
Qy 593 VRYNQSADFYGAGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 652
Db 593 VRYNQSADFYGAGEIYFKPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRPFFIAQD 652
Qy 653 DONAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRR 712
Db 653 DONAPRVPAAARLGFHLKASLTDRIDANLDYRVFAQNKLARYETRTPGHMLNLGANYRR 712
Qy 713 NTRYGEWNWYVKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758
Db 713 NTRYGEWNWYVKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF 758

RESULT 5
AAU74565
ID AAU74565 standard; Protein; 758 AA.
XX AAU74565;
AC AAU74565;
XX 21-MAR-2000 (first entry)
DT 21-MAR-2000 (first entry)
XX *Neisseria meningitidis* ORF 149 protein sequence SEQ ID NO:604.
DE *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX *Neisseria meningitidis*.
OS *Neisseria meningitidis*.
XX WO9957280-A2.
FN 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
PD 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.

```
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
DR N-PSDB; AAZ53327.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX Claim 2; Page 422-423; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 758 AA;
XX
XX Query Match 25.3%; Score 192; DB 21; Length 758;
XX Best Local Similarity 100.0%; Pred. No. 4.4e-192;
XX Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 38 VVGKSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDGVPGIHSQVGGASAPVI 97
XX Db ///////////////////////////////////////////////////////////////////
XX 38 VVGKSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDGVPGIHSQVGGASAPVI 97
XX
XX QY 98 RGGTGRIKVLNHHGETGMDADSPDHAIMVDTALSOQVEILRGPTVLLYSNGVAGLVD 157
XX Db ///////////////////////////////////////////////////////////////////
XX 98 RGGTGRIKVLNHHGETGMDADSPDHAIMVDTALSOQVEILRGPTVLLYSNGVAGLVD 157
XX
XX QY 158 VADGKIPEKMPENGVSSELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYKRSQDYAV 217
XX Db ///////////////////////////////////////////////////////////////////
XX 158 VADGKIPEKMPENGVSSELGLRLSSGNLEKLTSGGINIGLGNFVLHTEGLYKRSQDYAV 217
XX
XX QY 218 PRYNLRKLDPDS 229
XX Db ///////////////////////////////////////////////////////////////////
XX 218 PRYNLRKLDPDS 229
XX
XX RESULT 6
XX AAZ74563
XX ID AAZ74563 standard; Protein; 339 AA.
XX
XX AC AAZ74563;
XX
XX 21-MAR-2000 (first entry)
XX
XX DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:600.
XX
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy.
XX
XX OS Neisseria meningitidis.
XX
```

```
PN WO9957280-A2.
XX
XX PD 11-NOV-1999.
XX
XX PF 30-APR-1999; 99WO-US09346.
XX
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tettelin H, Venter JC;
XX
XX DR WPI: 2000-062150/05.
XX
XX DR N-PSDB; AAZ53325.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX
XX Claim 2; Page 420; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 339 AA;
XX
XX Query Match 23.1%; Score 175; DB 21; Length 339;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-174;
XX Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 473 PLPDLGAHROTARSAFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 532
XX Db ///////////////////////////////////////////////////////////////////
XX 54 PLPDLGAHROTARSAFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 113
XX
XX QY 533 NKHLNKRNSNNIELALGYEGDRWQYNLALYRNFPGNYIYAQTLDNGRGPXSIEDDSMKL 592
XX Db ///////////////////////////////////////////////////////////////////
XX 114 NKHLNKRNSNNIELALGYEGDRWQYNLALYRNFPGNYIYAQTLDNGRGPXSIEDDSMKL 173
XX
XX QY 593 VRYNQSGADFYGAEGRIYFKPTPRYRIGVSGDVYVRGLKNLPSLPGRDAYGNRP 647
XX Db ///////////////////////////////////////////////////////////////////
XX 174 VRYNQSGADFYGAEGRIYFKPTPRYRIGVSGDVYVRGLKNLPSLPGRDAYGNRP 228
XX
XX RESULT 7
XX AAZ74557
XX ID AAZ74557 standard; Protein; 764 AA.
XX
XX AC AAZ74557;
XX
XX 21-MAR-2000 (first entry)
XX
XX DE Neisseria meningitidis ORF 147 protein sequence SEQ ID NO:588.
XX
```

KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.

XX *Neisseria meningitidis*.

FN WO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

DR N-PSDB; AAZ53319.

XX Novel *Neisseria* polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

XX Claim 2; Page 414-415; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941

CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides

CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent

CC PCR primers used in the exemplification of the present invention. The

CC polypeptides, the polynucleotides, antibodies and compositions of

CC the invention can be used as vaccines, as diagnostic reagents, and as

CC immunogenic compositions. The polypeptides can be used in the

CC manufacture of medicaments for treating or preventing infection due to

CC *Neisseria* bacteria (e.g. meningitis and septicaemia), to detect the

CC presence of *Neisseria* bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves

CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.

XX Sequence 764 AA;

Query Match 23.1%; Score 175; DB 21; Length 764;

Best Local Similarity 100.0%; Pred. No. 3.1e-174;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 PLPDGAHQRTARFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 532

DB 479 PLPDGAHQRTARFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 538

AC ABB78068;

XX 05-NOV-2002 (first entry)

XX Amino acid sequence of p88 polypeptide.

XX p177; p88; p64; p55; p46; vaccine; gonorrhea.

XX *Neisseria gonorrhoeae*.

XX WO200260936-A2.

XX 08-AUG-2002.

XX 31-JAN-2002; 2002WO-US02881.

PR 31-JAN-2001; 2001US-266070P.

PR 06-AUG-2001; 2001US-310356P.

PR 23-OCT-2001; 2001US-344452P.

XX (IOWA) UNIV IOWA RES FOUND.

PA (REGC) UNIV CALIFORNIA.

PA (APIC/) APICELLA M. A.

PA (EDWA/) EDWARDS J. L.

PA (GIES/) GIBSON B. W.

PA (SCHE/) SCHEFFLER K.

PA (BROW/) BROWN E.

XX Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;

XX WPI; 2002-619227/66.

DR N-PSDB; ABQ78299.

XX New polypeptide comprising p177, p88, p64, p55 or p46 from *Neisseria*

PT *gonorrhoeae*, useful for preventing, or protecting a female patient

PT against, *N. gonorrhoeae* colonization or infection -

XX Claim 7; Page 115-117; 130pp; English.

XX The present sequence represents a p88 polypeptide. The specification

CC describes p177, p88, p64, p55 and p46 polypeptides from *Neisseria*

CC *gonorrhoeae*. The polypeptides are useful as vaccines, for preventing,

CC or protecting a female patient against, *N. gonorrhoeae* colonization or

CC infection. Such immunisation can prevent gonorrhea in women.

XX Sequence 764 AA;

Query Match 23.1%; Score 175; DB 23; Length 764;

Best Local Similarity 100.0%; Pred. No. 3.1e-174;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 PLPDGAHQRTARFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 532

DB 479 PLPDGAHQRTARFALSGNWFYTPQHKLSLTASHQERLPSTQELYAHGKHVATNTFEVG 538

QY 533 NKHLKERSNNIELALGYEGDRWQYNLALYRNRFNGNYIYAQTLDNGRGPKSIEDDSEMKL 592

DB 539 NKHLKERSNNIELALGYEGDRWQYNLALYRNRFNGNYIYAQTLDNGRGPKSIEDDSEMKL 598

QY 593 VRYNQSGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRP 647

DB 599 VRYNQSGADFYGAGEIYFKETPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRP 653

RESULT 9

ABP80923

ID ABP80923 standard; Protein; 728 AA.

XX ABB78068

XX 07-MAR-2003 (first entry)

XX N. gonorrhoeae amino acid sequence SEQ ID 8376.

XX Antibacterial; infection; vaccine; gene therapy.
 KW Neisseria gonorrhoeae.
 OS WO200279243-A2.
 XX 10-OCT-2002.
 XX 12-FEB-2002; 2002WO-IB02069.
 XX 12-FEB-2001; 2001GB-0003424.
 XX (CHIR-) CHIRON SPA.
 XX Fontana MR, Pizza M, Masignani V, Monaci E;
 XX WPI; 2003-058415/05.
 XX N-PSDB; ABZ41893.
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection -
 XX Disclosure; Page 799; 815pp; English.
 XX The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 XX
 SQ Sequence 728 AA;
 Query Match 18.2%; Score 138; DB 24; Length 728;
 Best Local Similarity 100.0%; Pred. No. 2.2e-135;
 Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 VVGKSRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGDLGVPGIHASQYGGASAPVI 97
 DB 8 VVGKSRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGDLGVPGIHASQYGGASAPVI 67
 QY 98 RGQTGRRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPTVLLYSSGNVAGLVD 157
 DB 68 RGQTGRRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPTVLLYSSGNVAGLVD 127
 QY 158 VADGKIPEKMPENGVSIGE 175
 DB 128 VADGKIPEKMPENGVSIGE 145
 RESULT 10
 AAY74564
 ID AAY74564 standard; Protein; 758 AA.
 XX
 AC AAY74564;
 XX
 XX 21-MAR-2000 (first entry)
 XX
 DE Neisseria gonorrhoeae ORF 149 protein sequence SEQ ID NO:602.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 XX
 XX WO9957280-A2.
 XX 11-NOV-1999.
 XX

PF 30-APR-1999; 99WO-US09346.
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 XX WPI; 2000-062150/05.
 DR N-PSDB; AAZ53326.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 421-422; 1453pp; English.
 XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 758 AA;
 Query Match 18.2%; Score 138; DB 21; Length 758;
 Best Local Similarity 100.0%; Pred. No. 2.3e-135;
 Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 38 VVGKSRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGDLGVPGIHASQYGGASAPVI 97
 DB 38 VVGKSRPRATSGLLHTSTASDKIIISGDTLRQKAVNLGDLGVPGIHASQYGGASAPVI 97
 QY 98 RGQTGRRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPTVLLYSSGNVAGLVD 157
 DB 98 RGQTGRRIKVLNHHGETGDMADFPDHAIMVDTALSQQVEILRGPTVLLYSSGNVAGLVD 157
 QY 158 VADGKIPEKMPENGVSIGE 175
 DB 158 VADGKIPEKMPENGVSIGE 175
 RESULT 11
 AAY74566
 ID AAY74566 standard; Protein; 758 AA.
 XX
 AC AAY74566;
 XX
 XX 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 149 protein sequence SEQ ID NO:606.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy.
 XX

```

OS Neisseria meningitidis.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53328.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 2; Page 424-425; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AA74253 to AA75941
XX represent novel Neisseria meningitis and N. gonorrhea polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX Sequence 758 AA;
XX Query Match 17.9%; Score 136; DB 21; Length 758;
XX Best Local Similarity 100.0%; Pred. No. 2.8e-133;
XX Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 473 PLPLDGAHROTARSPALSGNWFYTPQHKLSLTASHQERLPSQELIYAHGKHVATNTFVG 532
Db 473 PLPLDGAHROTARSPALSGNWFYTPQHKLSLTASHQERLPSQELIYAHGKHVATNTFVG 532
QY 533 NKHLNKERSNNIELALGYEGDRWYNLALYRNRFNGYIYAOTLNDGRGPKSIEDSEMKL 592
Db 533 NKHLNKERSNNIELALGYEGDRWYNLALYRNRFNGYIYAOTLNDGRGPKSIEDSEMKL 592
QY 593 VRYNQSGADFYGAEGE 608
Db 593 VRYNQSGADFYGAEGE 608
RESULT 12
AA74555
ID AA74555 standard; Protein; 174 AA.
XX AC AA74555;
XX AC AA74555;
DT 21-MAR-2000 (first entry)
XX AC AA74551;

```

```

DE Neisseria gonorrhea ORF 147 protein sequence SEQ ID NO:584.
XX Neisseria meningitidis; Neisseria gonorrhea; antigen; vaccine;
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX antibacterial; gene therapy.
XX OS Neisseria gonorrhea.
XX WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
XX N-PSDB; AAZ53317.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX Claim 2; Page 412; 1453pp; English.
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AA74253 to AA75941
XX represent novel Neisseria meningitis and N. gonorrhea polynucleotides
XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX Sequence 174 AA;
XX Query Match 15.4%; Score 117; DB 21; Length 174;
XX Best Local Similarity 100.0%; Pred. No. 6.3e-114;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 VVGKSRPRATSGLLHTSTASDKIISGDTLRKAVNLGDLGVPGIHASQYGGGASAFVI 97
Db 44 VVGKSRPRATSGLLHTSTASDKIISGDTLRKAVNLGDLGVPGIHASQYGGGASAFVI 103
QY 98 RGQTRRIKVLNHHGETGDMADFPDHALIWDVTALSSQVEILRGFVILLYSSGNVAG 154
Db 104 RGQTRRIKVLNHHGETGDMADFPDHALIWDVTALSSQVEILRGFVILLYSSGNVAG 160
RESULT 13
AA74561
ID AA74561 standard; Protein; 339 AA.
XX AC AA74561;
XX AC AA74561;

```

```

DT 21-MAR-2000 (first entry)
XX Neisseria gonorrhoeae ORF 149 protein sequence SEQ ID NO:596.
DE
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHTR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
DR N-PSDB; AA253323.
XX
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX
PS Claim 2; Page 418; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 339 AA;
Query Match 10.6%; Score 80; DB 21; Length 339;
Best Local Similarity 100.0%; Pred. No. 8.6e-75;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 562 YNRFGNYIAQTLDNGRGPKEIEDSEMKLVRYNQSGADFYGAGEIFYFKPTPRYRIGV 621
Db 143 YNRFGNYIAQTLDNGRGPKEIEDSEMKLVRYNQSGADFYGAGEIFYFKPTPRYRIGV 202
QY 622 SGDYVRGRLKNLPSLPGRED 641
Db 203 SGDYVRGRLKNLPSLPGRED 222
RESULT 14
AAG15254
ID AAG15254 standard; Protein; 144 AA.
XX

```

```

AC AAG15254;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 15434.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.

```

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
Db 9 ALSATSEAV 17
|||||

RESULT 15
AAG15253
ID AAG15253 standard; Protein; 145 AA.
XX
AC AAG15253;
XX
XX 17-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 15433.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW

KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0126785.

PR 06-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 20-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0135629.

PR 27-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136392.

PR 01-JUN-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137222.

PR 04-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137502.

PR 08-JUN-1999; 99US-0137724.

PR 10-JUN-1999; 99US-0138094.

PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149175.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

PR	07-SEP-1999;	99US-0152363.	PD	06-SEP-2000.	
PR	10-SEP-1999;	99US-0153070.	XX		
PR	13-SEP-1999;	99US-0153758.	PF	25-FEB-2000;	2000EP-0301439.
PR	15-SEP-1999;	99US-0154018.	XX		
PR	16-SEP-1999;	99US-0154039.	XX		
PR	20-SEP-1999;	99US-0154779.	PR	25-FEB-1999;	99US-0121825.
PR	22-SEP-1999;	99US-0155139.	PR	05-MAR-1999;	99US-0123180.
PR	23-SEP-1999;	99US-0155486.	PR	03-MAR-1999;	99US-0123548.
PR	24-SEP-1999;	99US-0155659.	PR	23-MAR-1999;	99US-0125788.
PR	28-SEP-1999;	99US-0156458.	PR	25-MAR-1999;	99US-0126264.
PR	29-SEP-1999;	99US-0156596.	PR	29-MAR-1999;	99US-0126785.
PR	04-OCT-1999;	99US-0157117.	PR	01-APR-1999;	99US-0127462.
PR	05-OCT-1999;	99US-0157753.	PR	06-APR-1999;	99US-0128234.
PR	06-OCT-1999;	99US-0157865.	PR	08-APR-1999;	99US-0128714.
PR	07-OCT-1999;	99US-0158029.	PR	16-APR-1999;	99US-0129845.
PR	08-OCT-1999;	99US-0158232.	PR	19-APR-1999;	99US-0130077.
PR	12-OCT-1999;	99US-0158369.	PR	21-APR-1999;	99US-0130449.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0131449.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159330.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159331.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159637.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999;	99US-0132486.
PR	18-OCT-1999;	99US-0159584.	PR	06-MAY-1999;	99US-0132487.
PR	21-OCT-1999;	99US-0160741.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160767.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160815.	PR	14-MAY-1999;	99US-0134370.
PR	22-OCT-1999;	99US-0160980.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160981.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160989.	PR	20-MAY-1999;	99US-0135124.
PR	23-OCT-1999;	99US-0161404.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161405.	PR	21-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161406.	PR	25-MAY-1999;	99US-0136021.
PR	26-OCT-1999;	99US-0161359.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161360.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161361.	PR	01-JUN-1999;	99US-0137222.
PR	28-OCT-1999;	99US-0161920.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161992.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161993.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
Query Match 1.2%; Score 9; DB 21; Length 145;					
Best Local Similarity 100.0%; Pred. No. 1.4;					
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	408	ALSATSEAV 416			
Db	10	ALSATSEAV 18			
RESULT 16					
AAG15252					
ID	AAG15252 standard; Protein; 151 AA.				
XX					
AC	AAG15252;				
XX					
DT	17-OCT-2000 (first entry)				
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 15432.				
XX					
KW	Protein identification; signal transduction pathway; metabolic pathway;				
KW	hybridisation assay; genetic mapping; gene expression control; promoter;				
XX	termination sequence.				
OS	Arabidopsis thaliana.				
XX					
PN	EP1033405-A2.				
XX					

```
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147240.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148585.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 408 ALSATSEAV 416
Db 16 ALSATSEAV 24

RESULT 17
AAG07631
ID AAG07631 standard; Protein; 323 AA.
XX AAG07631;
AC
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 4857.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
```

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159233.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132486.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999;	99US-0132487.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160741.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160767.	PR	11-MAY-1999;	99US-0132456.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160815.	PR	14-MAY-1999;	99US-0134370.
PR	22-OCT-1999;	99US-0160980.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160981.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160989.	PR	20-MAY-1999;	99US-0135124.
PR	25-OCT-1999;	99US-0161404.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161405.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161406.	PR	25-MAY-1999;	99US-0136021.
PR	26-OCT-1999;	99US-0161359.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161360.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161361.	PR	01-JUN-1999;	99US-0137222.
PR	28-OCT-1999;	99US-0161920.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161922.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161993.	PR	07-JUN-1999;	99US-0137724.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUN-1999;	99US-0138094.
			PR	10-JUN-1999;	99US-0138540.
			PR	10-JUN-1999;	99US-0138847.
			PR	14-JUN-1999;	99US-0139119.
			PR	16-JUN-1999;	99US-0139452.
			PR	16-JUN-1999;	99US-0139453.
			PR	17-JUN-1999;	99US-0139452.
			PR	18-JUN-1999;	99US-0139454.
			PR	18-JUN-1999;	99US-0139455.
			PR	18-JUN-1999;	99US-0139456.
			PR	18-JUN-1999;	99US-0139457.
			PR	18-JUN-1999;	99US-0139457.
			PR	18-JUN-1999;	99US-0139458.
			PR	18-JUN-1999;	99US-0139459.
			PR	18-JUN-1999;	99US-0139460.
			PR	18-JUN-1999;	99US-0139461.
			PR	18-JUN-1999;	99US-0139462.
			PR	18-JUN-1999;	99US-0139463.
			PR	18-JUN-1999;	99US-0139750.
			PR	18-JUN-1999;	99US-0139763.
			PR	21-JUN-1999;	99US-0139817.
			PR	22-JUN-1999;	99US-0139899.
			PR	23-JUN-1999;	99US-0140353.
			PR	23-JUN-1999;	99US-0140354.
			PR	24-JUN-1999;	99US-0140695.
			PR	28-JUN-1999;	99US-0140823.
			PR	29-JUN-1999;	99US-0140991.
			PR	30-JUN-1999;	99US-0141287.
			PR	01-JUL-1999;	99US-0141842.
			PR	01-JUL-1999;	99US-0142154.
			PR	02-JUL-1999;	99US-0142055.
			PR	06-JUL-1999;	99US-0142390.
			PR	08-JUL-1999;	99US-0142803.
			PR	09-JUL-1999;	99US-0142920.
			PR	12-JUL-1999;	99US-0142977.
			PR	13-JUL-1999;	99US-0143542.
			PR	14-JUL-1999;	99US-0143624.
			PR	15-JUL-1999;	99US-0144005.
			PR	16-JUL-1999;	99US-0144085.
			PR	16-JUL-1999;	99US-0144086.
			PR	19-JUL-1999;	99US-0144325.
			PR	19-JUL-1999;	99US-0144331.
			PR	19-JUL-1999;	99US-0144332.
			PR	19-JUL-1999;	99US-0144333.
			PR	19-JUL-1999;	99US-0144334.

Query Match 1.2%; Score 9; DB 21; Length 323;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
|||||
Db 9 ALSATSEAV 17

RESULT 18
AAG36807
ID AAG36807 standard; Protein; 323 AA.
XX AC AAG36807;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 45159.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EF1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151920.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 323;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 408 ALSATSEAV 416
Db 9 ALSATSEAV 17

RESULT 19

ABB93042
ID ABB93042 standard; Protein; 323 AA.
XX AC ABB93042;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 2253.

XX Herbicidal; plant; agriculture; herbicide.

OS Arabidopsis thaliana.

PN W0200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weidler M;

DR WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -

PS Claim 5; SEQ ID NO 2253; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins

CC (ABB90790-ABB94016) for herbicidally active compounds, comprising

CC aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using

CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
XX useful as herbicides.

SQ Sequence 323 AA;

Query Match 1.2%; Score 9; DB 23; Length 323;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416

Db |||||
9 ALSATSEAV 17

RESULT 20

AAG07630

ID AAG07630 standard; Protein; 324 AA.

XX AC AAG07630;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 4856.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; Genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 03-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127482.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 08-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

```
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 324;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
Db 10 ALSATSEAV 18
|||||
|

RESULT 21
AAG36806
ID AAG36806 standard; Protein; 324 AA.
XX
AC AAG36806;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45158.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0128785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
```


PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142977.
 PR 12-JUL-1999; 99US-0143542.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145313.
 PR 27-JUL-1999; 99US-0145318.
 PR 27-JUL-1999; 99US-0145319.
 PR 28-JUL-1999; 99US-0145391.
 PR 28-JUL-1999; 99US-0145391.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147303.
 PR 09-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149829.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 16-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 25-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161933.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 324;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 408 ALSATSEAV 416
 |||||

Db 10 ALSATSEAV 18

RESULT 22
AAG07629
ID AAG07629 standard; Protein; 330 AA.
XX AAG07629;
AC AAG07629;
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 4855.
DE Arabidopsis thaliana
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136399.
PR 28-MAY-1999; 99US-0136789.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147304.
PR 05-AUG-1999; 99US-0147152.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

PR	18-AUG-1999;	99US-0149426.	AC	AAG36805;
PR	20-AUG-1999;	99US-0149723.	XX	18-OCT-2000 (first entry)
PR	20-AUG-1999;	99US-0149723.	DT	Arabidopsis thaliana protein fragment SEQ ID NO: 45157.
PR	20-AUG-1999;	99US-0149929.	XX	Protein identification; signal transduction pathway; metabolic pathway;
PR	23-AUG-1999;	99US-0149902.	DE	hybridisation assay; genetic mapping; gene expression control; promoter;
PR	23-AUG-1999;	99US-0149930.	XX	termination sequence.
PR	25-AUG-1999;	99US-0150566.	KW	Arabidopsis thaliana.
PR	25-AUG-1999;	99US-0150884.	KW	EPI033405-A2.
PR	26-AUG-1999;	99US-0151065.	KW	06-SEP-2000.
PR	27-AUG-1999;	99US-0151065.	XX	25-FEB-2000; 2000EP-0301439.
PR	27-AUG-1999;	99US-0151066.	XX	25-FEB-1999; 99US-0121825.
PR	27-AUG-1999;	99US-0151080.	XX	05-MAR-1999; 99US-0123180.
PR	30-AUG-1999;	99US-0151303.	XX	09-MAR-1999; 99US-0123548.
PR	31-AUG-1999;	99US-0151438.	XX	23-MAR-1999; 99US-0125788.
PR	01-SEP-1999;	99US-0151930.	XX	25-MAR-1999; 99US-0126264.
PR	07-SEP-1999;	99US-0152363.	XX	29-MAR-1999; 99US-0126785.
PR	10-SEP-1999;	99US-0153070.	XX	01-APR-1999; 99US-0127462.
PR	13-SEP-1999;	99US-0153758.	XX	06-APR-1999; 99US-0128234.
PR	15-SEP-1999;	99US-0154018.	XX	08-APR-1999; 99US-0128714.
PR	16-SEP-1999;	99US-0154039.	XX	16-APR-1999; 99US-0129845.
PR	20-SEP-1999;	99US-0154779.	XX	19-APR-1999; 99US-0130077.
PR	22-SEP-1999;	99US-0155139.	XX	21-APR-1999; 99US-0130449.
PR	23-SEP-1999;	99US-0155486.	XX	23-APR-1999; 99US-0130510.
PR	24-SEP-1999;	99US-0155659.	XX	28-APR-1999; 99US-0130891.
PR	24-SEP-1999;	99US-0156458.	XX	30-APR-1999; 99US-0131449.
PR	28-SEP-1999;	99US-0156596.	XX	30-APR-1999; 99US-0132048.
PR	04-OCT-1999;	99US-0157117.	XX	04-MAY-1999; 99US-0132407.
PR	05-OCT-1999;	99US-0157753.	XX	05-MAY-1999; 99US-0132485.
PR	06-OCT-1999;	99US-0157865.	XX	06-MAY-1999; 99US-0132487.
PR	07-OCT-1999;	99US-0158029.	XX	07-MAY-1999; 99US-0132863.
PR	08-OCT-1999;	99US-0158232.	XX	11-MAY-1999; 99US-0134256.
PR	12-OCT-1999;	99US-0158369.	XX	14-MAY-1999; 99US-0134218.
PR	13-OCT-1999;	99US-0159293.	XX	14-MAY-1999; 99US-0134219.
PR	13-OCT-1999;	99US-0159294.	XX	14-MAY-1999; 99US-0134221.
PR	14-OCT-1999;	99US-0159329.	XX	18-MAY-1999; 99US-0134768.
PR	14-OCT-1999;	99US-0159330.	XX	19-MAY-1999; 99US-0134941.
PR	14-OCT-1999;	99US-0159331.	XX	20-MAY-1999; 99US-0135124.
PR	14-OCT-1999;	99US-0159637.	XX	21-MAY-1999; 99US-0135353.
PR	14-OCT-1999;	99US-0159638.	XX	24-MAY-1999; 99US-0135629.
PR	18-OCT-1999;	99US-0159584.	XX	25-MAY-1999; 99US-0136021.
PR	21-OCT-1999;	99US-0160741.	XX	27-MAY-1999; 99US-0136392.
PR	21-OCT-1999;	99US-0160767.	XX	28-MAY-1999; 99US-0136782.
PR	21-OCT-1999;	99US-0160768.	XX	01-JUN-1999; 99US-0137222.
PR	21-OCT-1999;	99US-0160770.	XX	03-JUN-1999; 99US-0137528.
PR	21-OCT-1999;	99US-0160814.	XX	04-JUN-1999; 99US-0137502.
PR	21-OCT-1999;	99US-0160815.	XX	07-JUN-1999; 99US-0137724.
PR	22-OCT-1999;	99US-0160980.	XX	08-JUN-1999; 99US-0138094.
PR	22-OCT-1999;	99US-0160981.	XX	10-JUN-1999; 99US-0138540.
PR	22-OCT-1999;	99US-0160981.	XX	10-JUN-1999; 99US-0138847.
PR	25-OCT-1999;	99US-0161404.	XX	14-JUN-1999; 99US-0139119.
PR	25-OCT-1999;	99US-0161405.	XX	16-JUN-1999; 99US-0139452.
PR	25-OCT-1999;	99US-0161406.	XX	16-JUN-1999; 99US-0139453.
PR	26-OCT-1999;	99US-0161359.	XX	17-JUN-1999; 99US-0139492.
PR	26-OCT-1999;	99US-0161360.	XX	18-JUN-1999; 99US-0139454.
PR	26-OCT-1999;	99US-0161361.	XX	18-JUN-1999; 99US-0139455.
PR	28-OCT-1999;	99US-0161920.	XX	18-JUN-1999; 99US-0139456.
PR	28-OCT-1999;	99US-0161920.	XX	18-JUN-1999; 99US-0139457.
PR	28-OCT-1999;	99US-0161992.	XX	18-JUN-1999; 99US-0139458.
PR	28-OCT-1999;	99US-0161993.	XX	18-JUN-1999; 99US-0139459.
PR	29-OCT-1999;	99US-0162142.	XX	18-JUN-1999; 99US-0139460.
PR	29-OCT-1999;	99US-0162142.	XX	18-JUN-1999; 99US-0139461.
PR	29-OCT-1999;	99US-0162142.	XX	18-JUN-1999; 99US-0139462.

Query Match 1.24; Score 9; DB 21; Length 330;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

408 ALSATSEAV 416
16 ALSATSEAV 24

RESULT 23
AAG36805
ID AAG36805 standard; Protein; 330 AA.
XX

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159323.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 9; DB 21; Length 330;
Best Local Similarity 100.0%; Pred. No. 3.1;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 ALSATSEAV 416
| | | | |
Db 16 ALSATSEAV 24

RESULT 24
ABR41659
ID ABR41659 standard; Protein; 1240 AA.
XX
AC ABR41659;
XX
XX 02-JUN-2003 (first entry)
XX Human DITHP cytoskeletal protein.
DE Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW

KW cancer; cell proliferative disorder; autoimmune disorder;
 KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
 KW neurological disorder; gastrointestinal disorder; transport disorder;
 KW connective tissue disorder; drug screening; proteome analysis;
 KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
 KW disease model; toxicological testing; transcript imaging;
 KW cytoskeletal protein.
 XX
 XX Homo sapiens.
 OS
 XX W0200297031-A2.
 PN
 XX
 PD 05-DEC-2002.
 XX
 XX 27-MAR-2002; 2002WO-US10056.
 PF
 XX 28-MAR-2001; 2001US-279619P.
 PR 29-MAR-2001; 2001US-280067P.
 PR 29-MAR-2001; 2001US-280068P.
 PR 16-MAY-2001; 2001US-291280P.
 PR 17-MAY-2001; 2001US-291829P.
 PR 17-MAY-2001; 2001US-291849P.
 PR 18-JUN-2001; 2001US-299428P.
 PR 20-JUN-2001; 2001US-299776P.
 PR 20-JUN-2001; 2001US-300001P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 PI Dfourth GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshley SR;
 PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
 XX
 XX WPI: 2003-129518/12.
 DR
 DR N-PSDB; ACC46596.
 DR
 XX
 PT Novel human diagnostic and therapeutic polypeptide useful for
 PT identifying test compound which specifically binds to a polypeptide
 PT encoded by human diagnostic and therapeutic polynucleotide, and to
 PT induce antibodies
 XX
 PS Claim 27; SEQ ID No 1194; 591pp; English.
 PS
 CC The invention relates to novel human diagnostic and therapeutic
 CC polynucleotides designated dithp (ACC46080-ACC46749) and to their
 CC encoded proteins (DITHP; ABR41136-ABR41812). The invention also relates
 CC to polynucleotide sequences at least 90% identical to the dithp cDNA
 CC sequences of the invention; recombinant vectors, host cells and
 CC transgenic organisms comprising a dithp nucleic acid sequence; the
 CC recombinant production of DITHP proteins; antibodies specific for DITHP
 CC proteins; microarrays comprising dithp nucleic acid sequences; methods
 CC for detecting dithp nucleotide and protein sequences; methods of screening
 CC for compounds which specifically bind a DITHP protein; and methods of
 CC assessing the toxicity of test compounds using a dithp hybridisation
 CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
 CC diagnosis of a wide variety of conditions including cancer and other cell
 CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
 CC viral, fungal or parasitic infections; hormonal disorders; metabolic
 CC disorders; neurological disorders; gastrointestinal disorders; transport
 CC disorders; and connective tissue disorders. They may also be used to
 CC screen for modulators of protein activity or gene expression. DITHP
 CC proteins can additionally be used in analysis of the proteome of a tissue
 CC or cell type and to induce antibodies. The dithp nucleic acids are
 CC additionally useful in somatic or germ-line gene therapy of the disorders
 CC mentioned above, as a source of antisense sequences, as a source of
 CC probes and primers, in genotyping and identification of individuals, in
 CC the generation of transgenic animal models of human disease or knock in
 CC humanised animals, in toxicological testing, and in transcript imaging.
 CC The present sequence represents a DITHP protein which is a cytoskeletal
 CC protein.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

```

ID  ABB47432 standard; Protein; 99 AA.
XX  ABB47432;
AC
XX
XX  05-FEB-2002 (first entry)
DT
XX
XX  Listeria monocytogenes protein #136.
DE
XX
XX  Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW  vitamin B12; bacterial infection; disease.
XX
XX  Listeria monocytogenes.
OS
XX
XX  W0200177335-A2.
PN
XX
XX  18-OCT-2001.
PD
XX
XX  11-APR-2001; 2001WO-FR011118.
PF
XX
XX  11-APR-2000; 2000FR-0004629.
PR
XX
XX  (INSP ) INST PASTEUR.
PA
XX
XX  Buchrieser C, Frangeul L, Couve E, Rusniok C, Psihi H, Dehoux P;
PI  Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI  Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI  Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI  Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant N;
PI  Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI  Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI  Rose M, Voss H;
XX
XX  WPI; 2002-010914/01.
DR
XX
XX  Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT  and prevention of Listeria and related bacterial infections, and
PT  related polypeptides -
XX
XX  Claim 6; SEQ ID No 137; 192pp; French.
PS
XX
XX  The present invention relates to the genome sequence of Listeria
CC  monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC  it are useful for selecting probes and primers for detecting genes in L.
CC  monocytogenes and related organisms, and for studying genetic
CC  polymorphisms and other genomes. The present sequence is a protein
CC  encoded by the genome sequence of the present invention. Proteins
CC  expressed from the genome sequence are useful for raising specific
CC  antibodies, identification of L. monocytogenes and related organisms, and
CC  for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC  B12. The genome sequence and proteins encoded by it are also useful for
CC  selecting compounds that regulate gene expression and cell replication
CC  and modulate L. monocytogenes-related diseases. In addition, the genome
CC  sequence and proteins encoded by it are useful in pharmaceutical and
CC  vaccines compositions for the treatment or prevention of infections by L.
CC  monocytogenes and related organisms.
CC  Note: The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published\_pct\_sequences.
XX
XX  Sequence 99 AA;
SQ
Query Match 1.1%; Score 8; DB 23; Length 99;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AQTTLKPI 9
Db 46 AQTTLKPI 53
|||||
|||||

RESULT 27
ABP32719
ID ABP32719 standard; Protein; 115 AA.

```

```

XX  ABP32719;
AC
XX
XX  08-JUL-2002 (first entry)
DT
XX
XX  Human ORF1692 protein, SEQ ID NO:3384.
DE
XX
XX  Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW  disease monitoring; cytokine; cell proliferation; cell differentiation;
KW  immune modulation; haematopoiesis regulation; tissue growth;
KW  angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW  thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW  behaviour; cancer; proliferative disorder; neurological disorder;
KW  cardiovascular disease; immune system disorder; organ transplantation;
KW  tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW  hypothyroidism; cholesterol ester storage disease; infection; vulnery;
KW  vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW  neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW  cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW  dermatological; analgesic; virucide; antibacterial; fungicide.
XX
XX  Homo sapiens.
OS
XX
XX  W0200190366-A2.
PN
XX
XX  29-NOV-2001.
PD
XX
XX  24-MAY-2001; 2001WO-US17076.
PF
XX
XX  24-MAY-2000; 2000US-206690P.
PR
XX
XX  (CURA-) CURAGEN CORP.
PA
XX
XX  Leach MD, Shimmets RA;
PI
XX
XX  WPI; 2002-106200/14.
DR
XX
XX  N-PSDB; ABN76745.
DR
XX
XX  Novel human polypeptides and polynucleotides useful for diagnosing,
PT  preventing and treating cardiovascular disease, neurodegenerative,
PT  hyperproliferative disorders and disorders related to organ
PT  transplantation -
XX
XX  Claim 10; Page 1105; 2508pp; English.
PS
XX
XX  Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC  designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC  ABN79587 represent cDNAs encoding them. The invention also encompasses
CC  polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC  referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC  the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC  polynucleotides, the recombinant production of ORFX proteins, antibodies
CC  specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC  polypeptides, methods of screening for modulators of ORFX expression or
CC  activity, and methods of screening individuals for a predisposition to an
CC  ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC  range of biological activities, such as cytokine, cell proliferation,
CC  cell differentiation, immune modulation, haematopoiesis regulation,
CC  tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC  chemokinetic activity, haemostatic activity, thrombolytic activity,
CC  receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC  and antiinfective activity, and may also be involved in the determination
CC  of bodily characteristics, fertility and behaviour. ORFX proteins,
CC  nucleic acids and antibodies may be used in the treatment of cancers,
CC  other proliferative disorders such as psoriasis and benign tumours,
CC  neurological disorders such as epilepsy and Alzheimer's disease,
CC  cardiovascular diseases, immune system disorders, disorders related to
CC  organ transplantation, disorders of tissue growth and regeneration,
CC  diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC  storage disease, and infectious diseases caused by viral, bacterial,
CC  fungal and other pathogens. ORFX nucleic acids may also be used as a
CC  source of primers and probes, in the detection of ORFX genomic sequences
CC  or transcripts, in the identification and cloning of homologous

```

CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.
 XX
 SQ Sequence 115 AA;

Query Match 1.1%; Score 8; DB 23; Length 115;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 EGLYRKSG 213
 DB 96 EGLYRKSG 103

RESULT 28
 ABB57898
 ID ABB57898 standard; Protein; 268 AA.
 XX
 AC ABB57898;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 486.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEXE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL02001.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX
 PS Disclosure; SEQ ID NO 486; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 268 AA;

Query Match 1.1%; Score 8; DB 22; Length 268;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 540 RSNNIELA 547
 DB 67 RSNNIELA 74
 RESULT 29
 AAU69473
 ID AAU69473 standard; Protein; 313 AA.
 XX
 AC AAU69473;
 XX
 DT 29-JAN-2002 (first entry)
 XX
 DE Human purified secretory polypeptide #42.
 XX
 KW Human; purified secretory polypeptide; cell proliferative disorder;
 KW cancer; immune system disorder; neurological disorder; mental disorder;
 KW motor neuron disorder; demyelinating disorder; neuromuscular disorder;
 KW central nervous system disorder; enzyme linked immunosorbent assay;
 KW ELISA; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200162918-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 01-FEB-2001; 2001WO-US03465.
 XX
 PR 24-FEB-2000; 2000US-185215P.
 PR 24-FEB-2000; 2000US-185216P.
 PR 16-MAY-2000; 2000US-205232P.
 PR 17-MAY-2000; 2000US-205286P.
 PR 17-MAY-2000; 2000US-205287P.
 PR 17-MAY-2000; 2000US-205323P.
 PR 17-MAY-2000; 2000US-205324P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'sa SA, Amehy S, Dahl CR, Dam TC, Daniels SE;
 PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo PD, Stockdreher TK, Daffo A;
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
 PI Cohen HJ, Hodgson DM, Lincoln SE;
 XX
 DR WPI; 2001-648217/74.
 DR
 XX
 PT Nucleic acids encoding secretory polypeptides, useful in genetic
 PT diagnosis and therapy -
 XX
 PS Disclosure; Page 214-215; 237pp; English.

CC Sequences AAU69432-AAU69511 represent purified secretory polypeptides of
 CC the invention. The polypeptides and their associated polynucleotides can
 CC be used in the treatment, prevention and diagnosis of diseases associated
 CC with inappropriate secretory protein expression. These diseases include
 CC cell proliferative disorders such as atherosclerosis and psoriasis,
 CC cancers such as leukaemia and melanoma, immune system disorders such as
 CC asthma and diabetes mellitus, neurological disorders such as epilepsy and
 CC Parkinson's disease, mental disorders such as schizophrenia and seasonal
 CC affective disorder (SAD), motor neuron disorders such as amyotrophic
 CC lateral sclerosis, demyelinating disorders such as multiple sclerosis,
 CC central nervous system disorders such as mental retardation and
 CC neurofibromatosis and neuromuscular disorders such as cerebral palsy and
 CC muscular dystrophy. Target polynucleotides in a sample can be detected by
 CC hybridising the sample with a probe sequence complementary to the target
 CC polynucleotide, under conditions in which a hybridisation complex is
 CC formed, and detecting the presence or absence of the complex. The
 CC polypeptides may also be used as antigens in the production of antibodies
 CC against secretory proteins and in assays to identify modulators of
 CC protein expression and activity. The antibodies may also be used as
 CC diagnostic agents for detecting the presence of the sequences of the

```
CC invention in samples e.g. by enzyme linked immunosorbent assay (ELISA).
SQ Sequence 313 AA;

Query Match 1.1%; Score 8; DB 22; Length 313;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 RVPAAARLG 665
Db 55 RVPAAARLG 62

RESULT 30
AAU99107
ID AAU99107 standard; Protein; 399 AA.
XX AC
XX AC
XX AAY85650;
XX DT 08-FEB-2001 (first entry)
XX DE Formate dehydrogenase amino acid sequence.
XX KW Formate dehydrogenase; Hyphomicrobium.
XX OS Hyphomicrobium sp.
XX PN JP2000245471-A.
XX PD 12-SEP-2000.
XX PF 01-MAR-1999; 99JP-0052548.
XX PR 01-MAR-1999; 99JP-0052548.
XX PA (NIRA ) UNITIKA LTD.
XX DR WPI; 2000-622119/60.
XX DR N-PSDB; AAC61183.
XX CC New Hyphomicrobium sp. formate dehydrogenase gene for producing formate
PT dehydrogenase of high specific activity, high temperature stability and
XX high pH stability
XX PS Claim 1; Page 5-6; 9pp; Japanese.
XX CC This invention relates to a gene encoding a formate dehydrogenase from
CC Hyphomicrobium sp. A transformant containing the new gene is used for the
CC preparation of formate dehydrogenase of high specific activity, low
CC Michaelis constant (Km) value against formic acid and NAD+, high
CC temperature stability and high pH stability. The present sequence
CC represents the protein of the invention.
XX SQ Sequence 399 AA;

Query Match 1.1%; Score 8; DB 21; Length 399;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
Db 54 VSGELGLR 61

RESULT 31
AAU99107
ID AAU99107 standard; Protein; 401 AA.
XX AC
XX AAU99107;
XX DT 24-SEP-2002 (first entry)
XX DE Mycobacterium vaccae formate dehydrogenase C6S/C146S/C256S mutant.
```

```
XX Formate dehydrogenase; alcohol production; ketone;
KW mutant; mutein.
XX Mycobacterium vaccae.
OS Synthetic.
XX FH Key Location/Qualifiers
FT Misc-difference 6 /note= "Wild type Cys substituted by Ser"
FT Misc-difference 146 /note= "Wild type Cys substituted by Ser"
FT Misc-difference 236 /note= "Wild type Cys substituted by Ser"
XX EP1211316-A1.
XX 05-JUN-2002.
XX PD
XX PF 27-NOV-2001; 2001EP-0128170.
XX PR 29-NOV-2000; 2000JP-0363894.
XX PR 24-AUG-2001; 2001JP-0254631.
XX PA (DAIL ) DAICEL CHEM IND LTD.
XX PI Mitsuhashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX DR
XX PT New mutant forms of formate dehydrogenase derived from Mycobacterium
PT vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials
XX Claim 8; Page -; 42pp; English.
XX CC This invention relates to a mutant form of the Mycobacterium vaccae
CC formate dehydrogenase protein which has strong formate dehydrogenase
CC activity in the presence of an organic solvent. The mutant formate
CC dehydrogenase protein of the invention may be used to produce alcohols
CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
CC proteins, the mutant polypeptides of the invention retain high enzymatic
CC activity in the presence of the alcohol product and so provides a higher
CC product yield. The present sequence represents a Mycobacterium vaccae
CC formate dehydrogenase protein mutant of the invention.
XX SQ Sequence 401 AA;

Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
Db 54 VSGELGLR 61

RESULT 32
AAU99108
ID AAU99108 standard; Protein; 401 AA.
XX AC AAU99108;
XX DT
XX DT 24-SEP-2002 (first entry)
XX DE Mycobacterium vaccae formate dehydrogenase C6A/C256S mutant.
XX KW Formate dehydrogenase; alcohol production; ketone;
XX KW mutant; mutein.
XX OS Mycobacterium vaccae.
OS Synthetic.
```



```

FH Key Location/Qualifiers
FT Misc-difference 6 /note= "Wild type Cys substituted by Ala"
FT Misc-difference 256
FT Misc-difference /note= "Wild type Cys substituted by Ser"
PN EP1211316-Al.
XX
XX 05-JUN-2002.
XX
XX 27-NOV-2001; 2001EP-0128170.
XX
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX (DAIL ) DAICEL CHEM IND LTD.
XX
XX Mitsuhashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX
XX New mutant forms of formate dehydrogenase derived from Mycobacterium
XX vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials -
XX
XX Claim 8; Page -; 42pp; English.
XX
XX This invention relates to a mutant form of the Mycobacterium vaccae
XX formate dehydrogenase protein which has strong formate dehydrogenase
XX activity in the presence of an organic solvent. The mutant formate
XX dehydrogenase protein of the invention may be used to produce alcohols
XX using ketones as raw materials. Unlike prior formatted dehydrogenase,
XX proteins, the mutant polypeptides of the invention retain high enzymatic
XX activity in the presence of the alcohol product and so provides a higher
XX product yield. The present sequence represents a Mycobacterium vaccae
XX formate dehydrogenase protein mutant of the invention.
XX
XX Sequence 401 AA;
XX
XX Query Match 1.1%; Score 8; DB 23; Length 401;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 172 VSIGELGLR 179
XX |||||
XX Db 54 VSIGELGLR 61
XX
XX RESULT 34
XX AAU99110
XX ID AAU99110 standard; Protein; 401 AA.
XX
XX AC AAU99110;
XX
XX DT 24-SEP-2002 (first entry)
XX
XX DE Mycobacterium vaccae formate dehydrogenase C6S/C256A mutant.
XX
XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.
XX
XX OS Mycobacterium vaccae.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 6 /note= "Wild type Cys substituted by Ser"
XX FT Misc-difference 256 /note= "Wild type Cys substituted by Ala"
XX FT Misc-difference 256 /note= "Wild type Cys substituted by Ala"
XX
XX PN EP1211316-Al.
XX
XX 05-JUN-2002.
XX
XX 27-NOV-2001; 2001EP-0128170.
XX
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX (DAIL ) DAICEL CHEM IND LTD.
XX
XX Mitsuhashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX
XX New mutant forms of formate dehydrogenase derived from Mycobacterium
XX vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials -
XX
XX Claim 8; Page -; 42pp; English.
XX
XX This invention relates to a mutant form of the Mycobacterium vaccae
XX formate dehydrogenase protein which has strong formate dehydrogenase
XX activity in the presence of an organic solvent. The mutant formate
XX dehydrogenase protein of the invention may be used to produce alcohols
XX using ketones as raw materials. Unlike prior formatted dehydrogenase,
XX proteins, the mutant polypeptides of the invention retain high enzymatic
XX activity in the presence of the alcohol product and so provides a higher
XX product yield. The present sequence represents a Mycobacterium vaccae
XX formate dehydrogenase protein mutant of the invention.
XX
XX Sequence 401 AA;
XX
XX Query Match 1.1%; Score 8; DB 23; Length 401;
XX Best Local Similarity 100.0%; Pred. No. 42;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 172 VSIGELGLR 179
XX |||||
XX Db 54 VSIGELGLR 61
XX
XX RESULT 33
XX AAU99109
XX ID AAU99109 standard; Protein; 401 AA.
XX
XX AC AAU99109;
XX
XX DT 24-SEP-2002 (first entry)
XX
XX DE Mycobacterium vaccae formate dehydrogenase C6V/C256S mutant.
XX
XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.
XX
XX OS Mycobacterium vaccae.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 6 /note= "Wild type Cys substituted by Val"
XX FT Misc-difference 256 /note= "Wild type Cys substituted by Ser"
XX FT Misc-difference 256 /note= "Wild type Cys substituted by Ser"
XX
XX PN EP1211316-Al.
XX
XX 05-JUN-2002.
XX

```

XX WPI; 2002-464925/50.
XX
PT New mutant forms of formate dehydrogenase derived from Mycobacterium
PT vaccae have enhanced activity in the presence of organic solvents and
PT are useful for producing alcohols from ketone raw materials -
XX
XX
PS Claim 8; Page -; 42pp; English.
XX
CC This invention relates to a mutant form of the Mycobacterium vaccae
CC formate dehydrogenase protein which has strong formate dehydrogenase
CC activity in the presence of an organic solvent. The mutant formate
CC dehydrogenase protein of the invention may be used to produce alcohols
CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
CC proteins, the mutant polypeptides of the invention retain high enzymatic
CC activity in the presence of the alcohol product and so provides a higher
CC product yield. The present sequence represents a Mycobacterium vaccae
CC formate dehydrogenase protein mutant of the invention.
XX
SQ Sequence 401 AA;
Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 172 VSGELGLR 179
Db 54 VSGELGLR 61
RESULT 35
AAU99111
ID AAU99111 standard; Protein; 401 AA.
AC AAU99111;
XX
XX 24-SEP-2002 (first entry)
XX Mycobacterium vaccae formate dehydrogenase C6S/C256V mutant.
XX
XX Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.
XX
XX Mycobacterium vaccae.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 6 /note= "Wild type Cys substituted by Ser"
FT Misc-difference 256
FT Misc-difference 256 /note= "Wild type Cys substituted by Val"
XX
XX EP1211316-A1.
XX
XX 05-JUN-2002.
XX
XX 27-NOV-2001; 2001EP-0128170.
XX
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX
XX (DAIL) DAICEL CHEM IND LTD.
XX
XX Mitsushashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX
XX 27-NOV-2001; 2001EP-0128170.
XX
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX
XX (DAIL) DAICEL CHEM IND LTD.
XX
XX Mitsushashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX
XX New mutant forms of formate dehydrogenase derived from Mycobacterium
XX vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials -
XX
XX Claim 8; Page -; 42pp; English.

CC This invention relates to a mutant form of the Mycobacterium vaccae
CC formate dehydrogenase protein which has strong formate dehydrogenase
CC activity in the presence of an organic solvent. The mutant formate
CC dehydrogenase protein of the invention may be used to produce alcohols
CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
CC proteins, the mutant polypeptides of the invention retain high enzymatic
CC activity in the presence of the alcohol product and so provides a higher
CC product yield. The present sequence represents a Mycobacterium vaccae
CC formate dehydrogenase protein mutant of the invention.
XX
SQ Sequence 401 AA;
Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 172 VSGELGLR 179
Db 54 VSGELGLR 61
RESULT 36
AAU99112
ID AAU99112 standard; Protein; 401 AA.
AC AAU99112;
XX
XX 24-SEP-2002 (first entry)
XX Mycobacterium vaccae formate dehydrogenase C146S mutant.
XX
XX Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.
XX
XX Mycobacterium vaccae.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 146 /note= "Wild type Cys substituted by Ser"
FT
XX
XX EP1211316-A1.
XX
XX 05-JUN-2002.
XX
XX 27-NOV-2001; 2001EP-0128170.
XX
XX 29-NOV-2000; 2000JP-0363894.
XX 24-AUG-2001; 2001JP-0254631.
XX
XX (DAIL) DAICEL CHEM IND LTD.
XX
XX Mitsushashi K, Yamamoto H, Kimoto N;
XX WPI; 2002-464925/50.
XX
XX New mutant forms of formate dehydrogenase derived from Mycobacterium
XX vaccae have enhanced activity in the presence of organic solvents and
XX are useful for producing alcohols from ketone raw materials -
XX
XX Claim 8; Page -; 42pp; English.
XX
CC This invention relates to a mutant form of the Mycobacterium vaccae
CC formate dehydrogenase protein which has strong formate dehydrogenase
CC activity in the presence of an organic solvent. The mutant formate
CC dehydrogenase protein of the invention may be used to produce alcohols
CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
CC proteins, the mutant polypeptides of the invention retain high enzymatic
CC activity in the presence of the alcohol product and so provides a higher
CC product yield. The present sequence represents a Mycobacterium vaccae
CC formate dehydrogenase protein mutant of the invention.
XX
SQ Sequence 401 AA;

Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
DB 54 VSGELGLR 61

RESULT 37

AAU99113
ID AAU99113 standard; Protein; 401 AA.

XX AC AAU99113;
XX

DT 24-SEP-2002 (first entry)

DE Mycobacterium vaccae formate dehydrogenase C256S mutant.

XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.

XX OS Mycobacterium vaccae.
XX Synthetic.

XX FH Key Location/Qualifiers
FT Misc-difference 256

FT /note= "Wild type Cys substituted by Ser"

XX EP1211316-A1.

XX OS 05-JUN-2002.

XX PF 27-NOV-2001; 2001EP-0128170.

XX PR 29-NOV-2000; 2000JP-0363894.

XX PR 24-AUG-2001; 2001JP-0254631.

XX PA (DAIL) DAICEL CHEM IND LTD.

XX PI Mitsuhashi K, Yamamoto H, Kimoto N;

XX WPI; 2002-464925/50.

XX New mutant forms of formate dehydrogenase derived from Mycobacterium
PT vaccae have enhanced activity in the presence of organic solvents and
PT are useful for producing alcohols from ketone raw materials

XX PS Claim 8; Page -; 42pp; English.

XX CC This invention relates to a mutant form of the Mycobacterium vaccae
CC formate dehydrogenase protein which has strong formate dehydrogenase
CC activity in the presence of an organic solvent. The mutant formate
CC dehydrogenase protein of the invention may be used to produce alcohols
CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
CC proteins, the mutant polypeptides of the invention retain high enzymatic
CC activity in the presence of the alcohol product and so provides a higher
CC product yield. The present sequence represents a Mycobacterium vaccae
CC formate dehydrogenase protein mutant of the invention.

XX SQ Sequence 401 AA;

Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
DB 54 VSGELGLR 61

RESULT 38

AAU99114
ID AAU99114 standard; Protein; 401 AA.

XX AC AAU99114;
XX

DT 24-SEP-2002 (first entry)

DE Mycobacterium vaccae formate dehydrogenase C146S/C256S mutant.

XX KW Formate dehydrogenase; alcohol production; ketone;
XX mutant; mutein.

XX OS Mycobacterium vaccae.
XX Synthetic.

XX FH Key Location/Qualifiers

FT Misc-difference 146

FT /note= "Wild type Cys substituted by Ser"

FT Misc-difference 256

FT /note= "Wild type Cys substituted by Ser"

XX EP1211316-A1.

XX PD 05-JUN-2002.

XX PF 27-NOV-2001; 2001EP-0128170.

XX PR 29-NOV-2000; 2000JP-0363894.

XX PR 24-AUG-2001; 2001JP-0254631.

XX PA (DAIL) DAICEL CHEM IND LTD.

XX PI Mitsuhashi K, Yamamoto H, Kimoto N;

XX WPI; 2002-464925/50.

XX New mutant forms of formate dehydrogenase derived from Mycobacterium
PT vaccae have enhanced activity in the presence of organic solvents and
PT are useful for producing alcohols from ketone raw materials

XX PS Claim 8; Page -; 42pp; English.

XX CC This invention relates to a mutant form of the Mycobacterium vaccae
CC formate dehydrogenase protein which has strong formate dehydrogenase
CC activity in the presence of an organic solvent. The mutant formate
CC dehydrogenase protein of the invention may be used to produce alcohols
CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
CC proteins, the mutant polypeptides of the invention retain high enzymatic
CC activity in the presence of the alcohol product and so provides a higher
CC product yield. The present sequence represents a Mycobacterium vaccae
CC formate dehydrogenase protein mutant of the invention.

XX SQ Sequence 401 AA;

Query Match 1.1%; Score 8; DB 23; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
DB 54 VSGELGLR 61

RESULT 39

AAU99115
ID AAU99115 standard; Protein; 401 AA.

XX AC AAU99115;
XX

DT 24-SEP-2002 (first entry)

DE Mycobacterium vaccae formate dehydrogenase C256V mutant.

XX

KW Formate dehydrogenase; alcohol production; ketone;
 KW mutant; mutin.
 XX
 OS Mycobacterium vaccae.
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 256 /note= "Wild type Cys substituted by Val"
 FT
 XX
 FN EP1211316-A1.
 XX
 PD 05-JUN-2002.
 XX
 PF 27-NOV-2001; 2001EP-0128170.
 XX
 PR 29-NOV-2000; 2000JP-0363894.
 PR 24-AUG-2001; 2001JP-0254631.
 XX
 PA (DAIL) DAICEL CHEM IND LTD.
 XX
 PI Mitsuhashi K, Yamamoto H, Kimoto N;
 XX
 DR WPI; 2002-464925/50.
 XX
 XX New mutant forms of formate dehydrogenase derived from Mycobacterium
 PT vaccae have enhanced activity in the presence of organic solvents and
 PT are useful for producing alcohols from ketone raw materials -
 XX
 PS Claim 8; Page -; 42pp; English.
 XX
 CC This invention relates to a mutant form of the Mycobacterium vaccae
 CC formate dehydrogenase protein which has strong formate dehydrogenase
 CC activity in the presence of an organic solvent. The mutant formate
 CC dehydrogenase protein of the invention may be used to produce alcohols
 CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
 CC proteins, the mutant polypeptides of the invention retain high enzymatic
 CC activity in the presence of the alcohol product and so provides a higher
 CC product yield. The present sequence represents a Mycobacterium vaccae
 CC formate dehydrogenase protein mutant of the invention.
 XX
 SQ Sequence 401 AA;
 Query Match 1.1%; Score 8; DB 23; Length 401;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 172 VSGELGLR 179
 Db 54 VSGELGLR 61
 RESULT 40
 AAU99116
 ID AAU99116 standard; Protein; 401 AA.
 XX
 AC AAU99116;
 XX
 DT 24-SEP-2002 (first entry)
 XX
 DE Mycobacterium vaccae formate dehydrogenase C146S/C256V mutant.
 XX
 KW Formate dehydrogenase; alcohol production; ketone;
 KW mutant; mutin.
 XX
 OS Mycobacterium vaccae.
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 146 /note= "Wild type Cys substituted by Ser"
 FT
 FT Misc-difference 256 /note= "Wild type Cys substituted by Val"

XX EP1211316-A1.
 XX
 PD 05-JUN-2002.
 XX
 PF 27-NOV-2001; 2001EP-0128170.
 XX
 PR 29-NOV-2000; 2000JP-0363894.
 PR 24-AUG-2001; 2001JP-0254631.
 XX
 PA (DAIL) DAICEL CHEM IND LTD.
 XX
 PI Mitsuhashi K, Yamamoto H, Kimoto N;
 XX
 DR WPI; 2002-464925/50.
 XX
 XX New mutant forms of formate dehydrogenase derived from Mycobacterium
 PT vaccae have enhanced activity in the presence of organic solvents and
 PT are useful for producing alcohols from ketone raw materials -
 XX
 PS Claim 8; Page -; 42pp; English.
 XX
 CC This invention relates to a mutant form of the Mycobacterium vaccae
 CC formate dehydrogenase protein which has strong formate dehydrogenase
 CC activity in the presence of an organic solvent. The mutant formate
 CC dehydrogenase protein of the invention may be used to produce alcohols
 CC using ketones as raw materials. Unlike prior formatted dehydrogenase,
 CC proteins, the mutant polypeptides of the invention retain high enzymatic
 CC activity in the presence of the alcohol product and so provides a higher
 CC product yield. The present sequence represents a Mycobacterium vaccae
 CC formate dehydrogenase protein mutant of the invention.
 XX
 SQ Sequence 401 AA;
 Query Match 1.1%; Score 8; DB 23; Length 401;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 172 VSGELGLR 179
 Db 54 VSGELGLR 61

Search completed: November 14, 2003, 11:06:54
 Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:06:59 ; Search time 21 Seconds
(without alignments)
1527.220 Million cell updates/sec

Title: US-09-936-377-2
Perfect score: 758
Sequence: 1 MAQTLXPIVLISILLINTPL.....FLSDTPQMRSGFTGVNVKF 758

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pdp.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pdp.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pdp.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pdp.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pdp.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.1	430	4	US-09-134-001C-2981
2	8	1.1	446	4	US-09-199-637A-267
3	8	1.1	704	4	US-09-328-352-5144
4	7	0.9	7	3	US-09-162-366C-18
5	7	0.9	8	3	US-09-162-366C-17
6	7	0.9	9	3	US-09-162-366C-16
7	7	0.9	10	3	US-09-162-366C-1
8	7	0.9	11	3	US-09-162-366C-4
9	7	0.9	11	3	US-09-162-366C-14
10	7	0.9	12	3	US-09-162-366C-12
11	7	0.9	13	3	US-09-162-366C-13
12	7	0.9	15	3	US-09-162-366C-11
13	7	0.9	18	3	US-09-162-366C-3
14	7	0.9	20	3	US-09-162-366C-2
15	7	0.9	51	4	US-09-205-258-963
16	7	0.9	112	3	US-08-928-361B-10
17	7	0.9	112	3	US-08-928-361B-29
18	7	0.9	112	4	US-09-588-995A-10
19	7	0.9	141	4	US-09-252-991A-23722
20	7	0.9	167	4	US-09-252-991A-30346
21	7	0.9	172	4	US-09-252-991A-20206
22	7	0.9	176	1	US-08-415-751-1
23	7	0.9	181	4	US-09-702-114A-15
24	7	0.9	218	4	US-09-252-991A-25202
25	7	0.9	230	4	US-09-252-991A-22090
26	7	0.9	235	4	US-09-252-991A-30274
27	7	0.9	243	4	US-09-489-847-205
28	7	0.9	247	4	US-09-252-991A-29972
29	7	0.9	258	4	US-09-134-001C-3244
30	7	0.9	261	2	US-07-857-224B-22
31	7	0.9	276	4	US-09-205-258-958
32	7	0.9	291	4	US-09-107-532A-3799
33	7	0.9	308	4	US-09-252-991A-24070
34	7	0.9	314	4	US-09-134-001C-2983
35	7	0.9	314	4	US-09-252-991A-25015
36	7	0.9	321	4	US-09-134-001C-2890
37	7	0.9	321	1	US-08-415-751-5
38	7	0.9	392	4	US-09-252-991A-25329
39	7	0.9	476	3	US-08-924-183-2
40	7	0.9	476	4	US-09-488-364-2
41	7	0.9	525	4	US-09-351-224E-8
42	7	0.9	525	4	US-09-677-682B-8
43	7	0.9	525	4	US-09-677-682B-30
44	7	0.9	562	4	US-08-851-567B-8
45	7	0.9	562	4	US-09-328-352-4594
46	7	0.9	569	4	US-09-252-991A-27248
47	7	0.9	611	4	US-09-107-532A-4988
48	7	0.9	688	4	US-09-252-991A-17098
49	7	0.9	740	4	US-09-252-991A-21187
50	7	0.9	789	4	US-09-252-991A-27011
51	7	0.9	1019	4	US-09-252-991A-24417
52	7	0.9	1046	4	US-09-252-991A-30284
53	7	0.9	1053	4	US-09-062-136-10
54	7	0.9	1189	4	US-08-851-567B-26
55	7	0.9	1311	2	US-08-540-406-4
56	7	0.9	1311	3	US-08-656-055-4
57	7	0.9	1311	3	US-08-954-688-4
58	7	0.9	1311	4	US-08-918-658-4
59	7	0.9	1311	4	US-09-724-631-4
60	7	0.9	1311	5	PCT-US95-13233-4
61	7	0.9	1721	3	US-08-700-651-5
62	7	0.9	1721	3	US-08-928-361B-6
63	7	0.9	1721	4	US-09-588-995A-5
64	7	0.9	1837	3	US-08-928-361B-5
65	7	0.9	1837	4	US-09-588-995A-5
66	7	0.9	2089	1	US-08-418-893D-23
67	7	0.9	2089	1	US-08-418-893D-24
68	7	0.9	3433	4	US-09-091-501B-10
69	7	0.9	4302	3	US-08-658-136-5
70	7	0.9	4302	4	US-09-052-469-8
71	7	0.9	4302	4	US-08-422-582-8
72	7	0.9	4303	2	US-08-460-751-2
73	7	0.9	4339	4	US-09-052-469-6
74	7	0.9	4339	4	US-08-422-582-6
75	6	0.8	6	3	US-08-665-643A-77
76	6	0.8	6	3	US-08-665-643A-78
77	6	0.8	6	3	US-08-665-643A-79
78	6	0.8	6	3	US-08-665-643A-80
79	6	0.8	6	3	US-08-665-643A-81
80	6	0.8	6	3	US-08-665-643A-82
81	6	0.8	6	3	US-08-665-643A-83
82	6	0.8	6	3	US-08-665-643A-84
83	6	0.8	6	3	US-08-665-643A-85
84	6	0.8	6	3	US-08-665-643A-86
85	6	0.8	6	3	US-08-665-643A-87
86	6	0.8	6	3	US-08-665-643A-88
87	6	0.8	6	3	US-08-665-643A-89
88	6	0.8	6	3	US-09-162-366C-19
89	6	0.8	6	3	US-08-665-643A-76
90	6	0.8	7	3	US-08-665-643A-43
91	6	0.8	7	4	US-09-640-198D-15
92	6	0.8	11	5	PCT-US91-09152-16
93	6	0.8	13	3	US-09-162-366C-10
94	6	0.8	19	4	US-09-161-939A-15
95	6	0.8	26	3	US-08-433-522A-47
96	6	0.8	26	3	US-09-135-166-47
97	6	0.8	26	3	US-08-942-046-47
98	6	0.8	28	1	US-08-427-072-11
99	6	0.8	30	4	US-09-205-258-929
100	6	0.8	31	2	US-08-210-762E-85

831	6	0.8	786	4	US-09-509-802-2	Sequence 2, Appli	904	6	0.8	969	5	PCT-US92-06888-3	Sequence 3, Appli
832	6	0.8	787	3	US-09-188-930-334	Sequence 334, App	905	6	0.8	969	5	PCT-US93-03027-4	Sequence 4, Appli
833	6	0.8	787	4	US-09-312-283C-334	Sequence 334, App	906	6	0.8	970	4	US-09-795-927-7	Sequence 7, Appli
834	6	0.8	790	4	US-09-634-957-2	Sequence 2, Appli	907	6	0.8	976	4	US-08-311-731A-363	Sequence 363, App
835	6	0.8	790	4	US-09-723-153-2	Sequence 2, Appli	908	6	0.8	989	3	US-09-110-517-4	Sequence 4, Appli
836	6	0.8	790	4	US-09-723-429-2	Sequence 2, Appli	909	6	0.8	990	4	US-09-252-991A-32469	Sequence 32469, A
837	6	0.8	792	4	US-09-252-991A-25823	Sequence 25823, A	910	6	0.8	994	4	US-09-521-511C-13	Sequence 13, Appli
838	6	0.8	793	4	US-09-523-849-32	Sequence 32, Appli	911	6	0.8	1007	4	US-09-252-991A-18614	Sequence 18614, A
839	6	0.8	797	3	US-08-433-522A-4	Sequence 4, Appli	912	6	0.8	1008	4	US-09-328-352-7403	Sequence 7403, Ap
840	6	0.8	797	3	US-08-433-522A-6	Sequence 4, Appli	913	6	0.8	1012	2	US-08-475-891A-4	Sequence 4, Appli
841	6	0.8	797	3	US-08-433-522A-6	Sequence 6, Appli	914	6	0.8	1025	2	US-08-567-375-4	Sequence 4, Appli
842	6	0.8	797	3	US-09-135-166-2	Sequence 2, Appli	915	6	0.8	1025	2	US-08-587-680A-4	Sequence 2, Appli
843	6	0.8	797	3	US-09-135-166-4	Sequence 4, Appli	916	6	0.8	1025	5	PCT-US93-04567-2	Sequence 2, Appli
844	6	0.8	797	3	US-09-135-166-6	Sequence 6, Appli	917	6	0.8	1037	4	US-09-134-001C-4794	Sequence 4794, Ap
845	6	0.8	797	3	US-08-942-046-2	Sequence 2, Appli	918	6	0.8	1044	4	US-09-252-991A-22493	Sequence 22493, A
846	6	0.8	797	3	US-08-942-046-4	Sequence 4, Appli	919	6	0.8	1044	4	US-09-107-532A-5229	Sequence 5229, Ap
847	6	0.8	797	3	US-08-942-046-6	Sequence 6, Appli	920	6	0.8	1065	3	US-08-630-172-9	Sequence 9, Appli
848	6	0.8	798	4	US-09-134-001C-4917	Sequence 4917, Ap	921	6	0.8	1065	3	US-09-375-419-9	Sequence 4, Appli
849	6	0.8	799	4	US-09-252-991A-26622	Sequence 26622, A	922	6	0.8	1074	1	US-07-912-015-4	Sequence 20, Appli
850	6	0.8	802	3	US-09-081-345-18	Sequence 18, Appli	923	6	0.8	1082	4	US-09-336-643A-20	Sequence 2, Appli
851	6	0.8	812	4	US-09-252-991A-30710	Sequence 30710, A	924	6	0.8	1083	4	US-09-600-776-2	Sequence 1, Appli
852	6	0.8	816	4	US-09-252-991A-23292	Sequence 23292, A	925	6	0.8	1083	4	US-09-343-494-1	Sequence 2, Appli
853	6	0.8	817	4	US-09-252-991A-27609	Sequence 27609, A	926	6	0.8	1083	4	US-09-358-383C-2	Sequence 2, Appli
854	6	0.8	827	4	US-09-252-991A-20443	Sequence 20443, A	927	6	0.8	1090	4	US-09-252-991A-23149	Sequence 23149, A
855	6	0.8	832	4	US-09-252-991A-22753	Sequence 22753, A	928	6	0.8	1104	4	US-09-252-991A-25463	Sequence 25463, A
856	6	0.8	833	4	US-09-514-302-3	Sequence 3, Appli	929	6	0.8	1123	4	US-09-252-991A-27976	Sequence 27976, A
857	6	0.8	836	4	US-09-417-197-61	Sequence 61, Appli	930	6	0.8	1132	4	US-09-198-452A-466	Sequence 466, App
858	6	0.8	837	4	US-09-122-126B-2	Sequence 2, Appli	931	6	0.8	1138	4	US-09-252-991A-20291	Sequence 2, Appli
859	6	0.8	837	4	US-09-634-286A-2	Sequence 2, Appli	932	6	0.8	1140	3	US-09-220-081-2	Sequence 2, Appli
860	6	0.8	842	4	US-09-417-197-43	Sequence 43, Appli	933	6	0.8	1140	4	US-09-677-575-2	Sequence 2, Appli
861	6	0.8	857	4	US-09-252-991A-31764	Sequence 31764, A	934	6	0.8	1148	2	US-08-313-185-58	Sequence 58, Appli
862	6	0.8	858	4	US-09-252-991A-25029	Sequence 25029, A	935	6	0.8	1148	3	US-09-082-614A-58	Sequence 58, Appli
863	6	0.8	861	4	US-09-784-316-2	Sequence 2, Appli	936	6	0.8	1148	3	US-09-106-075A-90	Sequence 90, Appli
864	6	0.8	862	4	US-09-252-991A-28455	Sequence 28455, A	937	6	0.8	1170	2	US-08-789-078-2	Sequence 2, Appli
865	6	0.8	864	4	US-08-714-741-40	Sequence 40, Appli	938	6	0.8	1170	2	US-08-752-633-2	Sequence 2, Appli
866	6	0.8	871	2	US-08-775-009-34	Sequence 34, Appli	939	6	0.8	1170	2	US-08-476-062A-42	Sequence 42, Appli
867	6	0.8	874	4	US-09-369-364A-15	Sequence 15, Appli	940	6	0.8	1170	5	PCT-US95-01314-42	Sequence 2, Appli
868	6	0.8	876	1	US-08-717-515-4	Sequence 4, Appli	941	6	0.8	1170	5	PCT-US96-01314-42	Sequence 2, Appli
869	6	0.8	876	2	US-08-510-215A-2	Sequence 2, Appli	942	6	0.8	1220	2	US-08-680-326-38	Sequence 38, Appli
870	6	0.8	876	4	US-09-517-871-20	Sequence 20, Appli	943	6	0.8	1220	4	US-08-107-322A-5150	Sequence 5150, Ap
871	6	0.8	876	4	US-09-252-991A-22193	Sequence 22193, A	944	6	0.8	1231	4	US-08-930-996A-4	Sequence 4, Appli
872	6	0.8	878	4	US-09-252-991A-31974	Sequence 31974, A	945	6	0.8	1231	3	US-08-930-996A-4	Sequence 74, Appli
873	6	0.8	879	4	US-09-914-259-38	Sequence 38, Appli	946	6	0.8	1269	3	US-08-781-891-74	Sequence 74, Appli
874	6	0.8	885	1	US-08-484-105-14	Sequence 14, Appli	947	6	0.8	1269	4	US-09-618-166-74	Sequence 74, Appli
875	6	0.8	885	1	US-08-484-106-14	Sequence 14, Appli	948	6	0.8	1277	1	US-08-717-515-8	Sequence 8, Appli
876	6	0.8	881	4	US-09-585-858-4	Sequence 4, Appli	949	6	0.8	1277	4	US-09-397-885-3	Sequence 3, Appli
877	6	0.8	882	4	US-09-585-858-2	Sequence 2, Appli	950	6	0.8	1277	4	US-09-969-362-3	Sequence 3, Appli
878	6	0.8	882	4	US-09-585-858-3	Sequence 3, Appli	951	6	0.8	1328	3	US-08-781-891-76	Sequence 76, Appli
879	6	0.8	882	4	US-09-585-858-3	Sequence 3, Appli	952	6	0.8	1328	4	US-09-618-166-76	Sequence 76, Appli
880	6	0.8	882	4	US-09-252-991A-32194	Sequence 32194, A	953	6	0.8	1375	3	US-09-171-410-1	Sequence 1, Appli
881	6	0.8	901	2	US-08-884-681-5	Sequence 5, Appli	954	6	0.8	1415	4	US-09-252-991A-26438	Sequence 26438, A
882	6	0.8	906	1	US-09-258-643-5	Sequence 5, Appli	955	6	0.8	1415	4	US-09-252-991A-31822	Sequence 31822, A
883	6	0.8	906	3	US-08-190-802A-31	Sequence 31, Appli	956	6	0.8	1423	3	US-08-810-712-10	Sequence 10, Appli
884	6	0.8	906	3	US-08-473-346-31	Sequence 31, Appli	957	6	0.8	1430	4	US-09-252-991A-18190	Sequence 4, Appli
885	6	0.8	906	4	US-08-473-089-31	Sequence 31, Appli	958	6	0.8	1451	1	US-08-308-872B-4	Sequence 22, Appli
886	6	0.8	911	3	US-08-487-072A-31	Sequence 31, Appli	959	6	0.8	1454	3	US-08-392-459-22	Sequence 26, Appli
887	6	0.8	915	4	US-08-460-269C-4	Sequence 4, Appli	960	6	0.8	1454	3	US-08-392-459-26	Sequence 32, Appli
888	6	0.8	922	3	US-08-460-269C-6	Sequence 6, Appli	961	6	0.8	1454	5	PCT-US91-08525-22	Sequence 22, Appli
889	6	0.8	923	4	US-09-397-885-1	Sequence 1, Appli	962	6	0.8	1454	5	PCT-US91-08525-26	Sequence 26, Appli
890	6	0.8	923	4	US-09-397-885-1	Sequence 1, Appli	963	6	0.8	1454	5	PCT-US91-08525-32	Sequence 32, Appli
891	6	0.8	934	1	US-08-215-805A-80	Sequence 80, Appli	964	6	0.8	1454	5	PCT-US93-04384-2	Sequence 2, Appli
892	6	0.8	934	3	US-08-929-329-6	Sequence 6, Appli	965	6	0.8	1454	5	PCT-US93-04384-8	Sequence 8, Appli
893	6	0.8	943	4	US-09-397-885-5	Sequence 5, Appli	966	6	0.8	1454	5	PCT-US93-04384-12	Sequence 12, Appli
894	6	0.8	943	4	US-09-397-885-5	Sequence 5, Appli	967	6	0.8	1454	5	PCT-US93-04384-16	Sequence 16, Appli
895	6	0.8	945	4	US-09-198-452A-112	Sequence 112, App	968	6	0.8	1454	5	PCT-US93-04384-43	Sequence 43, Appli
896	6	0.8	956	4	US-09-252-991A-17124	Sequence 17124, A	969	6	0.8	1454	5	PCT-US93-04384-44	Sequence 44, Appli
897	6	0.8	966	4	US-09-207-359B-47	Sequence 47, Appli	970	6	0.8	1454	5	PCT-US93-04384-45	Sequence 45, Appli
898	6	0.8	968	3	US-08-560-005-7	Sequence 7, Appli	971	6	0.8	1454	5	PCT-US93-04384-46	Sequence 46, Appli
899	6	0.8	968	3	US-09-418-540-7	Sequence 7, Appli	972	6	0.8	1454	5	PCT-US93-04384-47	Sequence 47, Appli
900	6	0.8	968	4	US-09-969-528-7	Sequence 7, Appli	973	6	0.8	1454	5	PCT-US93-04384-48	Sequence 48, Appli
901	6	0.8	969	1	US-08-365-689-3	Sequence 3, Appli	974	6	0.8	1477	1	US-08-038-682-4	Sequence 4, Appli
902	6	0.8	969	1	US-07-747-781-3	Sequence 3, Appli	975	6	0.8	1477	1	US-08-302-832-4	Sequence 4, Appli
903	6	0.8	969	1	US-08-145-138A-3	Sequence 3, Appli	976	6	0.8	1477	2	US-08-530-198-4	Sequence 4, Appli

```

977      6      0.8      1477      2      US-08-469-880-4      Sequence 4, Appli
978      6      0.8      1477      2      US-08-728-470-4      Sequence 4, Appli
979      6      0.8      1477      3      US-08-617-697-4      Sequence 10, Appli
980      6      0.8      1477      3      US-08-492-459-10      Sequence 10, Appli
981      6      0.8      1477      3      US-08-423-752-10      Sequence 3, Appli
982      6      0.8      1477      3      US-08-945-994-3      Sequence 24, Appli
983      6      0.8      1477      3      US-08-716-873-24      Sequence 4, Appli
984      6      0.8      1477      3      US-08-719-641-4      Sequence 24, Appli
985      6      0.8      1477      4      US-09-368-431-24      Sequence 10, Appli
986      6      0.8      1477      4      US-09-414-006-10      Sequence 10, Appli
987      6      0.8      1477      4      US-09-447-223-10      Sequence 71, Appli
988      6      0.8      1477      4      US-09-206-942-71      Sequence 5, Appli
989      6      0.8      1507      3      US-08-929-329-5      Patent No. 5268270
990      6      0.8      1507      6      5268270-2      Sequence 4, Appli
991      6      0.8      1514      2      US-08-853-310-4      Sequence 2, Appli
992      6      0.8      1536      1      US-08-038-682-2      Sequence 2, Appli
993      6      0.8      1536      1      US-08-302-832-2      Sequence 2, Appli
994      6      0.8      1536      2      US-08-530-198-2      Sequence 2, Appli
995      6      0.8      1536      2      US-08-469-880-2      Sequence 2, Appli
996      6      0.8      1536      2      US-08-728-470-2      Sequence 2, Appli
997      6      0.8      1536      2      US-08-617-697-2      Sequence 2, Appli
998      6      0.8      1536      3      US-08-719-641-2      Sequence 2, Appli
999      6      0.8      1536      4      US-09-206-942-67      Sequence 67, Appli
1000     6      0.8      1580      2      US-08-804-227C-11      Sequence 11, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-134-001C-2981
; Sequence 2981, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2981
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2981

```

```

Query Match      1.1%; Score 8; DB 4; Length 430;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      9      IVLSILLI 16
Db      231     IVLSILLI 238

```

```

RESULT 2
US-09-199-637A-267
; Sequence 267, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John

```

```

; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 80786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 267
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-267

Query Match      1.1%; Score 8; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      238     HRAVLGWR 245
Db      65      HRAVLGWR 72

```

```

RESULT 3
US-09-328-352-5144
; Sequence 5144, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5144
; LENGTH: 704
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5144

```

```

Query Match      1.1%; Score 8; DB 4; Length 704;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      256     RRDQYGLP 263
Db      267     RRDQYGLP 274

```

```

RESULT 4
US-09-162-366C-18
; Sequence 18, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18

```

```
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-18

Query Match          0.9%; Score 7; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 1 NLGDALD 7

RESULT 5
US-09-162-366C-17
; Sequence 17, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-17

Query Match          0.9%; Score 7; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 2 NLGDALD 8

RESULT 6
US-09-162-366C-16
; Sequence 16, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
```

```
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-16

Query Match          0.9%; Score 7; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.5e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 3 NLGDALD 9

RESULT 7
US-09-162-366C-1
; Sequence 1, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-1

Query Match          0.9%; Score 7; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 4 NLGDALD 10

RESULT 8
US-09-162-366C-4
; Sequence 4, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
```

```

; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-4

```

```

Query Match 0.9%; Score 7; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
Db 5 NLGDALD 11

```

```

RESULT 9
US-09-162-366C-14
; Sequence 14, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-14

```

```

Query Match 0.9%; Score 7; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
Db 5 NLGDALD 11

```

```

RESULT 10
US-09-162-366C-12
; Sequence 12, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil

```

```

; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-12

```

```

Query Match 0.9%; Score 7; DB 3; Length 12;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
Db 6 NLGDALD 12

```

```

RESULT 11
US-09-162-366C-13
; Sequence 13, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-13

```

```

Query Match 0.9%; Score 7; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 72 NLGDALD 78
Db 7 NLGDALD 13

```

```

RESULT 12
US-09-162-366C-11

```

; Sequence 11, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-11

Query Match 0.9%; Score 7; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14
|||||

RESULT 13
US-09-162-366C-3
; Sequence 3, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-3

Query Match 0.9%; Score 7; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14
|||||

Db 8 NLGDALD 14

RESULT 14
US-09-162-366C-2
; Sequence 2, Application US/09162366C
; Patent No. 6296850
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/162,366C
; CURRENT FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/060,556
; PRIOR FILING DATE: 1997-09-30
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-162-366C-2

Query Match 0.9%; Score 7; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 7 NLGDALD 13
|||||

RESULT 15
US-09-205-258-963
; Sequence 963, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007p1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894

```

; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 963
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-205-258-963

```

```

Query Match      0.9%; Score 7; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      153 AGLVDVA 159
      |||||

```

Db

14 AGLVDVA 20

RESULT 16

```

US-08-928-361B-10
; Sequence 10, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verity, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 112 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-10

```

```

Query Match      0.9%; Score 7; DB 3; Length 112;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy

4 TTLKPIV 10

Db

74 TTLKPIV 80

RESULT 17

```

US-08-928-361B-29
; Sequence 29, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto

```


STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928.361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verity, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 112 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-29

Query Match 0.9%; Score 7; DB 3; Length 112;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTLKPIV 10
Db 74 TTLKPIV 80

RESULT 18

US-09-588-995A-10
Sequence 10, Application US/09588995A
Patent No. 6514697
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: BARNES, DEBRA A.
APPLICANT: NELSON, RICHARD C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM INFECTIONS
FILE REFERENCE: 480.19-5
CURRENT APPLICATION NUMBER: US/09/588,995A
CURRENT FILING DATE: 2000-06-06
PRIOR APPLICATION NUMBER: 08/827,171
PRIOR FILING DATE: 1997-03-27
PRIOR APPLICATION NUMBER: 08/928,361
PRIOR FILING DATE: 1997-09-12
PRIOR APPLICATION NUMBER: 08/700,651
PRIOR FILING DATE: 1996-08-14
PRIOR APPLICATION NUMBER: 08/415,751
PRIOR FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 115
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 112
TYPE: PRT
ORGANISM: Cryptosporidium parvum
US-09-588-995A-10

Query Match 0.9%; Score 7; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 41;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 TTLKPIV 10
Db 74 TTLKPIV 80

RESULT 19

US-09-252-991A-23722
Sequence 23722, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23722
LENGTH: 141
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: UNSURE
LOCATION: (86)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-23722

Query Match 0.9%; Score 7; DB 4; Length 141;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 APRVPA 662
Db 13 APRVPA 19

RESULT 20

US-09-252-991A-30346
Sequence 30346, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30346
LENGTH: 167
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30346

Query Match 0.9%; Score 7; DB 4; Length 167;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 PRVPAAR 663
Db 112 PRVPAAR 118

RESULT 21

US-09-252-991A-20206
; Sequence 20206, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20206
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20206

Query Match 0.9%; Score 7; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 SSALSAT 412
|||||
Db 138 SSALSAT 144

RESULT 22

US-08-415-751-1
; Sequence 1, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS AND KIT
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PHILLIPS, MOORE, LEMPPIO & FINLEY
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: California
; COUNTRY: United States of America
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
; COMPUTER: PC
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA: US/08/415,751
; FILING DATE: 03-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,880
; FILING DATE: June 1, 1993
; APPLICATION NUMBER: 07/891,301
; FILING DATE: May 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hana Dolezalova
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 324-1677
; TELEFAX: (415) 324-1678
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Cryptosporidium parvum
US-08-415-751-1

Query Match 0.9%; Score 7; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTLKPIV 10
|||||
Db 69 TTLKPIV 75

RESULT 23

US-09-702-114A-15
; Sequence 15, Application US/09702114A
; Patent No. 6566078
; GENERAL INFORMATION:
; APPLICANT: Arthur B. Raitano
; APPLICANT: Aya Jakobovits
; APPLICANT: Mary Faris
; APPLICANT: Daniel E.H. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: 36P6D5: SECRETED TUMOR ANTIGEN
; FILE REFERENCE: 129.22-US-U1
; CURRENT APPLICATION NUMBER: US/09/702,114A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 60/162,417
; PRIOR FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-702-114A-15

Query Match 0.9%; Score 7; DB 4; Length 181;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 NGVSGEL 176
|||||
Db 67 NGVSGEL 73

RESULT 24

US-09-252-991A-25202
; Sequence 25202, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25202

```

; LENGTH: 218
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25202

Query Match      0.9%; Score 7; DB 4; Length 218;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 212 SGDYAVP 218
Db 125 SGDYAVP 131

RESULT 25
US-09-252-991A-22090
; Sequence 22090, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22090
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22090

Query Match      0.9%; Score 7; DB 4; Length 230;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 381 ARIELRH 387
Db 81 ARIELRH 87

RESULT 26
US-09-252-991A-30274
; Sequence 30274, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30274
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30274

Query Match      0.9%; Score 7; DB 4; Length 235;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 QHRAVLG 243
Db 237 QHRAVLG 243

```

```

Db 115 QHRAVLG 121

RESULT 27
US-09-489-847-205
; Sequence 205, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 205
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-489-847-205

Query Match      0.9%; Score 7; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 AGLVDVA 159
Db 206 AGLVDVA 212

RESULT 28
US-09-252-991A-29972
; Sequence 29972, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29972
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29972

Query Match      0.9%; Score 7; DB 4; Length 247;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 GGGASAP 95
Db 61 GGGASAP 67

```

```

RESULT 29
US-09-134-001C-3244
; Sequence 3244, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3244
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3244

Query Match      0.9%; Score 7; DB 4; Length 258;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      748 RSFTGGV 754
Db      236 RSFTGGV 242

RESULT 30
US-07-857-224B-22
; Sequence 22, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION:
; ORIGINAL SOURCE:
; ORGANISM: rabbit
; FEATURE: Protein kinase; Table 8 Column 24
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.

```

```

;
; AUTHORS: Quinn, A. M.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
US-07-857-224B-22

Query Match      0.9%; Score 7; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      489 LSGNWF 495
Db      213 LSGNWF 219

RESULT 31
US-09-205-258-958
; Sequence 958, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972

```

```

; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 958
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-203-258-958

Query Match      0.9%; Score 7; DB 4; Length 276;
Best Local Similarity 100.0%; Pred.No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLVDVA 159
Db 239 AGLVDVA 245

RESULT 32
US-09-532A-3799
; Sequence 3799, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3799:
SEQUENCE CHARACTERISTICS:
LENGTH: 291 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...291
SEQUENCE DESCRIPTION: SEQ ID NO: 3799:
US-09-107-532A-3799

Query Match      0.9%; Score 7; DB 4; Length 291;
Best Local Similarity 100.0%; Pred.No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ILLINTP 19
Db 240 ILLINTP 246

RESULT 33
US-09-252-991A-24070
; Sequence 24070, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24070
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24070

Query Match      0.9%; Score 7; DB 4; Length 308;
Best Local Similarity 100.0%; Pred.No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 EKAGDAV 370
Db 280 EKAGDAV 286
```

```
RESULT 34
US-09-134-001C-2983
; Sequence 2983, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2983
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2983

Query Match      0.9%; Score 7; DB 4; Length 314;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      73 LGDALDG 79
Db      300 LGDALDG 306

RESULT 35
US-09-252-991A-25015
; Sequence 25015, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25015
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25015

Query Match      0.9%; Score 7; DB 4; Length 314;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      476 DLGAHRQ 482
Db      138 DLGAHRQ 144

RESULT 36
US-09-134-001C-2890
; Sequence 2890, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
```

```
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2890
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2890

Query Match      0.9%; Score 7; DB 4; Length 321;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      751 TGGVNVK 757
Db      293 TGGVNVK 299

RESULT 37
US-08-415-751-5
; Sequence 5, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS AND KIT
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PHILLIPS, MOORE, LEMPFO & FINLEY
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: California
; COUNTRY: United States of America
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
; COMPUTER: PC
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,751
; FILING DATE: 03-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,880
; FILING DATE: June 1, 1993
; APPLICATION NUMBER: 07/891,301
; FILING DATE: May 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hana Dolezalova
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-1677
; TELEFAX: (415) 324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
```

;
; ORGANISM: Cryptosporidium parvum
; FEATURE:
; NAME/KEY: Positions coded by nonsense codons are
; NAME/KEY: identified as xaa.
US-08-415-751-5

Query Match 0.9%; Score 7; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTKKPIV 10
| | | | |
Db 151 TTKKPIV 157

RESULT 38
US-09-252-991A-25329
; Sequence 25329, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25329
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25329

Query Match 0.9%; Score 7; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 ALDGVPG 82
| | | | |
Db 184 ALDGVPG 190

RESULT 39
US-08-924-183-2
; Sequence 2, Application US/08924183A
; Patent No. 6218109
; GENERAL INFORMATION:
; APPLICANT: Ellledge, Stephen J.
; APPLICANT: Sanchez, Yolanda
; TITLE OF INVENTION: MAMMALIAN CHECKPOINT GENES AND PROTEINS
; FILE REFERENCE: 120541-1003
; CURRENT APPLICATION NUMBER: US/08/924,183A
; CURRENT FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-924-183-2

Query Match 0.9%; Score 7; DB 3; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 RPRATSG 49
| | | | |
Db 275 RPRATSG 281

RESULT 40
US-09-488-364-2
; Sequence 2, Application US/09488364
; Patent No. 6307015
; GENERAL INFORMATION:
; APPLICANT: Ellledge, Stephen J.
; APPLICANT: Sanchez, Yolanda
; TITLE OF INVENTION: MAMMALIAN CHECKPOINT GENES AND PROTEINS
; FILE REFERENCE: 120541-1013
; CURRENT APPLICATION NUMBER: US/09/488,364
; CURRENT FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-488-364-2

Query Match 0.9%; Score 7; DB 4; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 RPRATSG 49
| | | | |
Db 275 RPRATSG 281

Search completed: November 14, 2003, 11:09:15
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:08:49 ; Search time 35 Seconds
(without alignments)
3953.717 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 758

Sequence: 1 MAQTTLPKPVILSILLINPL.....FLSDTPQMGSRFTGGVNVKF 758

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	175	23.1	764	15	US-10-065-551-2
2	8	1.1	313	15	US-10-204-887-121
3	8	1.1	387	15	US-10-156-761-9479
4	8	1.1	401	12	US-09-996-008B-2
5	8	1.1	401	12	US-10-286-326-4
6	8	1.1	446	11	US-09-975-719-267
7	7	0.9	7	9	US-09-915-306-18
8	7	0.9	7	10	US-09-915-374-18
9	7	0.9	8	9	US-09-915-306-17
10	7	0.9	8	10	US-09-915-374-17
11	7	0.9	9	9	US-09-915-306-15
12	7	0.9	9	10	US-09-915-374-16
13	7	0.9	10	9	US-09-915-306-1
14	7	0.9	10	10	US-09-915-374-1
15	7	0.9	11	9	US-09-915-306-4
					Sequence 2, Appli
					Sequence 121, App
					Sequence 9479, Ap
					Sequence 2, Appli
					Sequence 4, Appli
					Sequence 267, App
					Sequence 18, Appl
					Sequence 17, Appl
					Sequence 17, Appl
					Sequence 16, Appl
					Sequence 16, Appl
					Sequence 1, Appli
					Sequence 1, Appli
					Sequence 4, Appli

11	9	US-09-915-306-14	Sequence 14, Appli
11	10	US-09-915-374-4	Sequence 4, Appli
11	10	US-09-915-374-14	Sequence 14, Appli
12	9	US-09-915-306-12	Sequence 12, Appli
12	10	US-09-915-374-12	Sequence 12, Appli
13	9	US-09-915-306-13	Sequence 13, Appli
13	10	US-09-915-374-13	Sequence 13, Appli
15	9	US-09-915-306-11	Sequence 11, Appli
15	9	US-09-915-374-11	Sequence 11, Appli
18	9	US-09-915-306-3	Sequence 3, Appli
18	10	US-09-915-374-3	Sequence 3, Appli
20	9	US-09-915-306-2	Sequence 2, Appli
20	10	US-09-915-374-2	Sequence 2, Appli
46	12	US-10-124-805-518	Sequence 518, App
46	14	US-10-007-805-518	Sequence 518, App
46	15	US-10-076-622-518	Sequence 518, App
47	9	US-09-864-761-42772	Sequence 42772, A
51	12	US-09-933-767-963	Sequence 963, App
51	15	US-10-023-282-963	Sequence 963, App
52	15	US-10-156-761-15079	Sequence 15079, A
111	15	US-10-156-761-7973	Sequence 7973, App
127	10	US-09-764-868-801	Sequence 801, App
136	12	US-10-123-058-6	Sequence 6, Appli
136	12	US-10-414-609-6	Sequence 6, Appli
147	15	US-10-153-668-364	Sequence 364, App
159	15	US-10-156-761-13330	Sequence 13330, A
210	15	US-10-037-340-133	Sequence 133, App
230	9	US-09-759-468-2	Sequence 2, Appli
232	10	US-09-919-497-78	Sequence 78, Appli
234	9	US-09-815-242-5314	Sequence 5314, Ap
241	9	US-09-815-242-12602	Sequence 12602, A
243	10	US-09-938-418-7	Sequence 7, Appli
243	11	US-09-946-374-352	Sequence 352, App
243	11	US-09-946-374-431	Sequence 431, App
243	12	US-10-015-387A-352	Sequence 352, App
243	12	US-10-015-387A-431	Sequence 431, App
243	12	US-10-063-735-122	Sequence 122, App
243	12	US-10-006-130A-352	Sequence 352, App
243	12	US-10-006-130A-431	Sequence 431, App
243	12	US-10-199-672-366	Sequence 366, App
243	12	US-10-006-172A-352	Sequence 352, App
243	12	US-10-006-172A-431	Sequence 431, App
243	12	US-10-187-749-366	Sequence 366, App
243	12	US-10-194-457-366	Sequence 366, App
243	12	US-10-184-642-366	Sequence 366, App
243	12	US-10-196-747-366	Sequence 366, App
243	12	US-10-015-392A-352	Sequence 352, App
243	12	US-10-015-392A-431	Sequence 431, App
243	12	US-10-017-253A-352	Sequence 352, App
243	12	US-10-017-253A-431	Sequence 431, App
243	12	US-10-124-805-514	Sequence 514, App
243	12	US-10-173-689-366	Sequence 366, App
243	12	US-10-173-690-366	Sequence 366, App
243	12	US-10-173-691-366	Sequence 366, App
243	12	US-10-173-692-366	Sequence 366, App
243	12	US-10-173-694-366	Sequence 366, App
243	12	US-10-173-698-366	Sequence 366, App
243	12	US-10-173-699-366	Sequence 366, App
243	12	US-10-173-707-366	Sequence 366, App
243	12	US-10-174-569-366	Sequence 366, App
243	12	US-10-174-583-366	Sequence 366, App
243	12	US-10-174-587-366	Sequence 366, App
243	12	US-10-174-591-366	Sequence 366, App
243	12	US-10-174-599-366	Sequence 366, App
243	12	US-10-175-732-366	Sequence 366, App
243	12	US-10-175-736-366	Sequence 366, App
243	12	US-10-175-742-366	Sequence 366, App
243	12	US-10-175-745-366	Sequence 366, App
243	12	US-10-175-748-366	Sequence 366, App
243	12	US-10-175-751-366	Sequence 366, App
243	12	US-10-175-754-366	Sequence 366, App
243	12	US-10-176-480-366	Sequence 366, App
243	12	US-10-176-489-366	Sequence 366, App

965 6 0.8 128 10 US-09-795-006A-71 Sequence 71, Appl
966 6 0.8 128 10 US-09-795-006A-73 Sequence 73, Appl
967 6 0.8 128 10 US-09-795-006A-75 Sequence 75, Appl
968 6 0.8 128 10 US-09-795-006A-79 Sequence 79, Appl
969 6 0.8 129 10 US-09-795-006A-49 Sequence 49, Appl
970 6 0.8 130 12 US-10-325-717-7 Sequence 7, Appl
971 6 0.8 130 14 US-10-072-159-11 Sequence 11, Appl
972 6 0.8 130 14 US-10-072-159-12 Sequence 12, Appl
973 6 0.8 130 14 US-10-072-159-13 Sequence 13, Appl
974 6 0.8 130 14 US-10-060-845-11 Sequence 11, Appl
975 6 0.8 130 14 US-10-060-845-12 Sequence 12, Appl
976 6 0.8 130 14 US-10-060-845-13 Sequence 13, Appl
977 6 0.8 133 15 US-10-012-542-409 Sequence 409, App
978 6 0.8 133 16 US-10-286-421-31 Sequence 31, Appl
979 6 0.8 135 12 US-10-325-717-110 Sequence 110, App
980 6 0.8 135 12 US-10-325-717-111 Sequence 111, App
981 6 0.8 135 12 US-10-325-717-113 Sequence 113, App
982 6 0.8 136 10 US-09-938-497-8 Sequence 8, Appl
983 6 0.8 136 10 US-09-938-497-14 Sequence 14, Appl
984 6 0.8 136 12 US-10-342-224-106 Sequence 106, App
985 6 0.8 137 12 US-10-029-586-30376 Sequence 30376, A
986 6 0.8 137 12 US-10-325-717-6 Sequence 6, Appl
987 6 0.8 138 11 US-09-847-208-155 Sequence 155, App
988 6 0.8 139 11 US-09-832-129-53 Sequence 53, Appl
989 6 0.8 139 11 US-09-272-975-51 Sequence 51, Appl
990 6 0.8 140 15 US-10-156-761-9549 Sequence 9549, App
991 6 0.8 142 15 US-10-227-616-103 Sequence 103, App
992 6 0.8 144 11 US-09-764-891-3838 Sequence 3838, App
993 6 0.8 144 12 US-10-325-717-5 Sequence 5, Appl
994 6 0.8 144 12 US-10-325-717-25 Sequence 25, Appl
995 6 0.8 144 12 US-10-325-717-41 Sequence 41, Appl
996 6 0.8 144 12 US-10-325-717-42 Sequence 42, Appl
997 6 0.8 144 12 US-10-325-717-43 Sequence 43, Appl
998 6 0.8 145 12 US-10-325-717-11 Sequence 11, Appl
999 6 0.8 145 12 US-10-325-717-27 Sequence 27, Appl
1000 6 0.8 145 12 US-10-325-717-28 Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-10-066-551-2
; Sequence 2, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; TITLE OF INVENTION: prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045U51
; CURRENT APPLICATION NUMBER: US/10/066,551
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-2

Query Match 23.1%; Score 175; DB 15; Length 764;
Best Local Similarity 100.0%; Pred. No. 2e-166;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 PLPLGHAHQRTARSAFALSGNMYFTPOHKLSLTASHQRLPSTQELYAHGKHVATNTFEVG 532
Db 479 PLPLGHAHQRTARSAFALSGNMYFTPOHKLSLTASHQRLPSTQELYAHGKHVATNTFEVG 538
Qy 533 NKHLNKRSSNNIELALGYEGDRWQYNLALYRNRTGNVIYAQTLDNDRGPKSIEDDSEMKL 592
Db 539 NKHLNKRSSNNIELALGYEGDRWQYNLALYRNRTGNVIYAQTLDNDRGPKSIEDDSEMKL 598
Qy 593 VRYNQSADFYGABGEIYFXTPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRP 647
Db 599 VRYNQSADFYGABGEIYFXTPTPRYRIGVSGDYVRGLKNLPSLPGRDAYGNRP 653

RESULT 2

US-10-204-887-121
; Sequence 121, Application US/10204887
; Publication No. US20030124569A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.
; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHEN, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan B.
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENAWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKREHER, Theresa K.
; APPLICANT: DARFO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1134 PCT
; CURRENT APPLICATION NUMBER: US/10/204,887
; CURRENT FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: 60/185,215
; PRIOR FILING DATE: 2000-02-24; 2000-05-16; 2000-05-17;
; PRIOR FILING DATE: 2000-02-24; 2000-05-16; 2000-05-17;
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PERL Program
; SEQ ID NO 121
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:1072337.2.orf2:2000MAY01
US-10-204-887-121

Query Match 1.1%; Score 8; DB 15; Length 313;
Best Local Similarity 100.0%; Pred. No. 34;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 RVPAARLG 665
|||||
Db 55 RVPAARLG 62

RESULT 3

US-10-156-761-9479
; Sequence 9479, Application US/10156761
; Publication No. US20030119018A1

GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9479
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9479

Query Match 1.1%; Score 8; DB 15; Length 387;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
|||||
Db 54 VSGELGLR 61

RESULT 4

US-09-996-008B-2
; Sequence 2, Application US/09996008B
; Publication No. US20030157677A1

GENERAL INFORMATION:

; APPLICANT: Mitsuhashi, Kazuya
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20030157677A1
; TITLE OF INVENTION: MUTANTS OF MYCOBACTERIUM VACCAR-DERIVED
; FILE REFERENCE: 14879-093001
; CURRENT APPLICATION NUMBER: US/09/996,008B
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: JP 2000-363894
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: JP 2001-254631
; PRIOR FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-996-008B-2

Query Match 1.1%; Score 8; DB 12; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179

Db 54 VSGELGLR 61
|||||

RESULT 5

US-10-286-326-4
; Sequence 4, Application US/10286326
; Publication No. US20030175903A1

GENERAL INFORMATION:

; APPLICANT: San, Ka-Yui
; APPLICANT: Berrios-Rivera, Susana
; APPLICANT: Bennett, George
; TITLE OF INVENTION: Recycling System for Manipulation of Intracellular NADH Availabi
; FILE REFERENCE: P02328U1
; CURRENT APPLICATION NUMBER: US/10/286,326
; CURRENT FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: US 60/335,371
; PRIOR FILING DATE: 2001-11-02
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Pseudomonas
US-10-286-326-4

Query Match 1.1%; Score 8; DB 12; Length 401;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VSGELGLR 179
|||||
Db 54 VSGELGLR 61

RESULT 6

US-09-975-719-267
; Sequence 267, Application US/09975719
; Publication No. US20030022349A1

GENERAL INFORMATION:

; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Rahme, Laurence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975,719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 267
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-267

Query Match 1.1%; Score 8; DB 11; Length 446;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 HRAVLGWR 245
|||||
Db 65 HRAVLGWR 72

RESULT 7

US-09-915-306-18
; Sequence 18, Application US/09915306
; Patent No. US20020081310A1

GENERAL INFORMATION:


```

; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-18

Query Match          0.9%; Score 7; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 72 NLGDALD 78
Db 1 NLGDALD 7

```

```

RESULT 8
US-09-915-374-18
; Sequence 18, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-18

Query Match          0.9%; Score 7; DB 10; Length 7;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 72 NLGDALD 78
Db 1 NLGDALD 7

```

```

RESULT 9
US-09-915-306-17
; Sequence 17, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-17

Query Match          0.9%; Score 7; DB 9; Length 8;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 72 NLGDALD 78
Db 2 NLGDALD 8

```

```

RESULT 10
US-09-915-374-17
; Sequence 17, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-17

Query Match          0.9%; Score 7; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 72 NLGDALD 78
Db 1 NLGDALD 7

```

QY 72 NLGDALD 78
Db 2 NLGDALD 8

RESULT 11

US-09-915-306-16
; Sequence 16, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-16

Query Match 0.9%; Score 7; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 3 NLGDALD 9

RESULT 12

US-09-915-374-16
; Sequence 16, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-16

Query Match 0.9%; Score 7; DB 10; Length 9;
Best Local Similarity 100.0%; Pred. No. 6e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 3 NLGDALD 9

RESULT 13

US-09-915-306-1
; Sequence 1, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-1

Query Match 0.9%; Score 7; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 4 NLGDALD 10

RESULT 14

US-09-915-374-1
; Sequence 1, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10
; TYPE: PRT

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-1

Query Match          0.9%; Score 7; DB 10; Length 10;
Best Local Similarity 100.0%; Pred.No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 NLGDALD 78
Db 4 NLGDALD 10

RESULT 15
US-09-915-306-4
; Sequence 4, Application US/09915306
; Patent No. US2002081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915.306
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-4

Query Match          0.9%; Score 7; DB 9; Length 11;
Best Local Similarity 100.0%; Pred.No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 NLGDALD 78
Db 5 NLGDALD 11

RESULT 16
US-09-915-306-14
; Sequence 14, Application US/09915306
; Patent No. US2002081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915.306
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
```

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-14

Query Match          0.9%; Score 7; DB 9; Length 11;
Best Local Similarity 100.0%; Pred.No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 NLGDALD 78
Db 5 NLGDALD 11

RESULT 17
US-09-915-374-4
; Sequence 4, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915.374
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-4

Query Match          0.9%; Score 7; DB 10; Length 11;
Best Local Similarity 100.0%; Pred.No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 NLGDALD 78
Db 5 NLGDALD 11

RESULT 18
US-09-915-374-14
; Sequence 14, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915.374
```

```
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 14
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-14

Query Match          0.9%; Score 7; DB 10; Length 11;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 5 NLGDALD 11

RESULT 19
US-09-915-306-12
; Sequence 12, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-12

Query Match          0.9%; Score 7; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 6 NLGDALD 12

RESULT 20
US-09-915-374-12
; Sequence 12, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
```

```
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-12

Query Match          0.9%; Score 7; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 6 NLGDALD 12

RESULT 21
US-09-915-306-13
; Sequence 13, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 13
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-13

Query Match          0.9%; Score 7; DB 9; Length 13;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 7 NLGDALD 13

RESULT 22
US-09-915-374-13
; Sequence 13, Application US/09915374
; Publication No. US20020197657A1
```

GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-13

Query Match 0.9%; Score 7; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 7 NLGDALD 13

RESULT 23
US-09-915-306-11
; Sequence 11, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-11

Query Match 0.9%; Score 7; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14

Query Match 0.9%; Score 7; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 24
US-09-915-374-11
; Sequence 11, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-11

Query Match 0.9%; Score 7; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 NLGDALD 78
Db 8 NLGDALD 14

RESULT 25
US-09-915-306-3
; Sequence 3, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 09/162,366
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-306-3

Query Match 0.9%; Score 7; DB 9; Length 18;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

US-09-915-306-2
Query Match      0.9%; Score 7; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 NLGDALD 78
Db      7 NLGDALD 13

RESULT 26
US-09-915-374-3
; Sequence 3, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-3

Query Match      0.9%; Score 7; DB 10; Length 18;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 NLGDALD 78
Db      8 NLGDALD 14

RESULT 27
US-09-915-306-2
; Sequence 2, Application US/09915306
; Patent No. US20020081310A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,306
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-2
Query Match      0.9%; Score 7; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 NLGDALD 78
Db      7 NLGDALD 13

RESULT 28
US-09-915-374-2
; Sequence 2, Application US/09915374
; Publication No. US20020197657A1
; GENERAL INFORMATION:
; APPLICANT: Bjorklund, Viveka
; APPLICANT: Bjorklund, Bertil
; APPLICANT: Bjorklund, Peter
; APPLICANT: Nap, Marius
; APPLICANT: Ramaekers, Frans C.S.
; APPLICANT: Schutte, Bert
; TITLE OF INVENTION: Apoptosis-Related Compounds and Their Use
; FILE REFERENCE: Bjorklund 09/162366
; CURRENT APPLICATION NUMBER: US/09/915,374
; CURRENT FILING DATE: 2001-07-27
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: SE 9703456-3
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: epitope from
; OTHER INFORMATION: mammalian protein
US-09-915-374-2

Query Match      0.9%; Score 7; DB 12; Length 46;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      72 NLGDALD 78
Db      7 NLGDALD 13

RESULT 29
US-10-124-805-518
; Sequence 518, Application US/10124805
; Publication No. US20030166022A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 518
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-124-805-518

Query Match      0.9%; Score 7; DB 12; Length 46;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLVDVA 159
 Db 9 AGLVDVA 15

RESULT 30

US-10-007-805-518
 ; Sequence 518, Application US/10007805
 ; Publication No. US20020150581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yuqiu
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Hepler, William T.
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: McNeill, Patricia D.
 ; APPLICANT: Durham, Margarita
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.470C10
 ; CURRENT APPLICATION NUMBER: US/10/007.805
 ; CURRENT FILING DATE: 2001-12-07
 ; NUMBER OF SEQ ID NOS: 593
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 518
 ; LENGTH: 46
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-007-805-518

Query Match 0.9%; Score 7; DB 14; Length 46;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLVDVA 159
 Db 9 AGLVDVA 15

RESULT 31

US-10-076-622-518
 ; Sequence 518, Application US/10076622
 ; Publication No. US20030023036A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Houghton, Raymond L.
 ; APPLICANT: Sleath, Paul R.
 ; APPLICANT: Persing, David H.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.470C11
 ; CURRENT APPLICATION NUMBER: US/10/076.622
 ; CURRENT FILING DATE: 2002-02-13
 ; NUMBER OF SEQ ID NOS: 627
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 518
 ; LENGTH: 46
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-076-622-518

Query Match 0.9%; Score 7; DB 15; Length 46;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLVDVA 159
 Db 9 AGLVDVA 15

RESULT 32

US-09-864-761-42772
 ; Sequence 42772, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecomica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864.761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 42772
 ; LENGTH: 47
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AL096678.8
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.97
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.89
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
 US-09-864-761-42772

Query Match 0.9%; Score 7; DB 9; Length 47;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

US-09-933-767-963
; Sequence 963, Application US/09933767
; Publication No. US20030181692A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P2
; CURRENT APPLICATION NUMBER: US/09/933,767
; PRIORITY FILING DATE: 2001-08-22
; PRIORITY APPLICATION NUMBER: PCT/US01/05614
; PRIORITY FILING DATE: 2001-02-21
; PRIORITY APPLICATION NUMBER: 60/184,836
; PRIORITY FILING DATE: 2000-02-24
; PRIORITY APPLICATION NUMBER: 60/193,170
; PRIORITY FILING DATE: 2000-03-29
; PRIORITY APPLICATION NUMBER: 09/205,258
; PRIORITY FILING DATE: 1998-12-04
; PRIORITY APPLICATION NUMBER: PCT/US98/11422
; PRIORITY FILING DATE: 1998-06-04
; PRIORITY APPLICATION NUMBER: 60/048,885
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/049,375
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,881
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,880
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,896
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/049,020
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,876
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,895
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,884
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,894
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,971
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,964
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,882
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,899
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,893
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,900
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,901
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,892
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,915
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/049,019
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,970
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,972
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/048,916
; PRIORITY FILING DATE: 1997-06-06
; PRIORITY APPLICATION NUMBER: 60/049,373
; PRIORITY FILING DATE: 1997-06-06

```



```

; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P1
; CURRENT APPLICATION NUMBER: US/10/023,282
; CURRENT FILING DATE: 2001-12-20
; EARLIER APPLICATION NUMBER: 09/205,258
; EARLIER FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898

```

```

; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 963
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-023-282-963

Query Match          0.9%; Score 7; DB 15; Length 51;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 AGLVDVA 159
DB 14 AGLVDVA 20
|||||

RESULT 35
US-10-156-761-15079
; Sequence 15079, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15079
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15079

Query Match          0.9%; Score 7; DB 15; Length 52;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 TLLYSSG 150
DB 7 TLLYSSG 13
|||||

RESULT 36
US-10-156-761-7973
; Sequence 7973, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

```

```
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7973
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7973

Query Match          0.9%; Score 7; DB 15; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 446 GGVREK 452
DB 28 GGVREK 34

RESULT 37
US-09-764-868-801
; Sequence 801, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND ANTIBODIES
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 801
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-801

Query Match          0.9%; Score 7; DB 10; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 489 LSGNWF 495
DB 26 LSGNWF 32

RESULT 38
US-10-123-058-6
; Sequence 6, Application US/10123058
; Publication No. US20030194695A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: MCGUIRE, MICHAEL J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES USEFUL AS
; TITLE OF INVENTION: ADJUVANTS
; FILE REFERENCE: UTSD:664US
; CURRENT APPLICATION NUMBER: US/10/123,058
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 6
; LENGTH: 136
; TYPE: PRT
; ORGANISM: PARAPOX
US-10-123-058-6

Query Match          0.9%; Score 7; DB 12; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 SRPRATS 48
DB 112 SRPRATS 118

RESULT 39
US-10-414-609-6
; Sequence 6, Application US/10414609
; Publication No. US20030194737A1
; GENERAL INFORMATION:
; APPLICANT: JOHNSTON, STEPHEN A.
; APPLICANT: MCGUIRE, MICHAEL J.
; TITLE OF INVENTION: Use of Parapox B2L Protein to Modify Immune Responses to Adminis
; TITLE OF INVENTION: Antigens
; FILE REFERENCE: UTSD:664-2US
; CURRENT APPLICATION NUMBER: US/10/414,609
; CURRENT FILING DATE: 2003-04-15
; PRIOR APPLICATION NUMBER: US 10/123,058
; PRIOR FILING DATE: 2002-04-15
; PRIOR APPLICATION NUMBER: US 60/336,694
; PRIOR FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 136
; TYPE: PRT
; ORGANISM: PARAPOX
US-10-414-609-6

Query Match          0.9%; Score 7; DB 12; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 SRPRATS 48
DB 112 SRPRATS 118

RESULT 40
US-10-153-668-364
; Sequence 364, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, GOICHI
; APPLICANT: MATSUDA, AKIO
; APPLICANT: MURAMATSU, SHUJI
; APPLICANT: ISHIZAWA, KENYA
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
```

```

; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 364
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-364

Query Match      0.9%; Score 7; DB 15; Length 147;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      495 FTPQKLL 501
Db      92 FTPQKLL 98

Search completed: November 14, 2003, 11:13:40
Job time : 37 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 16:20:31 ; Search time 4856 Seconds
(without alignments)
6385.811 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTILKPVLISILLINTPL.....FLSDTQMGSRFTGGVNVKF 758

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 288711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO_spoil/US9936377/runat_14112003_104449_22108/app_query.fasta_1.903
-DB=GenEmbl -OPMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US9936377_@CGN_1_1_3078_@runat_14112003_104449_22108 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DBEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.ov.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.in.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4036	100.0	2277	6	AX035174 Sequence
2	3928	97.3	2295	6	AX548017 Sequence
C 3	3928	97.3	331801	1	NMA422491
4	3927	97.3	2277	6	AX300375 Sequence
C 5	3927	97.3	10700	1	AE002448 Neisseria
C 6	3927	97.3	349980	6	AX044031 Sequence
7	1725.5	42.8	4931	1	AY028475 Sequence
8	1585	39.3	2457	6	AX052567 Sequence
9	1585	39.3	2646	6	AX394914 Sequence
10	1585	39.3	89047	6	AX067459 Sequence
C 11	1492.5	37.0	10120	1	AE006149 Pasteurel
C 12	1364	33.8	11375	1	AE006179 Pasteurel
C 13	1307	32.4	10166	1	AE006112 Pasteurel
14	1205.5	29.9	2942	6	AX394912 Sequence
15	1205.5	29.9	269223	6	AX067466 Sequence
16	1202	29.8	2742	6	AX052569 Sequence
17	745	18.5	10947	1	AE012378 Xanthomon
18	718.5	17.8	11947	1	AE011924 Xanthomon
19	694.5	17.2	10158	1	AE003890 Xylella f
C 20	691.5	17.1	295250	1	BX321862 Nitrosomo
C 21	689.5	17.1	300885	1	AE012559 Xylella f
C 22	679	16.8	15470	1	AE004513 Pseudomon
C 23	642	15.9	301995	1	AE016779 Pseudomon
C 24	636	15.8	11626	1	AE005695 Caulobact
C 25	469	11.6	300400	1	AP005943 Bradyrhiz
C 26	396	9.8	10214	1	AE011477 Bacteroid
C 27	392	9.7	303750	1	AE016931 Pasteurel
28	354.5	8.8	11549	1	AE006167 Pasteurel
29	290	7.2	17725	1	AE004349 Vibrio ch
30	271.5	6.7	1173	6	AX079015 Sequence
31	265	6.6	10161	1	AE005425 Escherich
32	265	6.6	286857	1	AP002559 Escherich
33	265	6.6	311143	1	AE016762 Escherich
34	264.5	6.6	2589	1	AF337562 Escherich
C 35	255	6.3	16841	1	AE004749 Pseudomon
C 36	253	6.3	10777	1	AE006093 Pasteurel
C 37	244.5	6.1	10399	1	AY008342 Plesiomon
38	244.5	6.1	12558	1	AB015838 Shewanell
C 39	243	6.0	16256	1	AE000694 Aquifex a
40	240.5	6.0	336182	1	AE016813 Vibrio vu
41	239	5.9	2501	1	AF127222 Pseudomon
42	238.5	5.9	7702	1	AF055999 Pseudomon
43	236.5	5.9	13223	1	AE004885 Pseudomon
44	236.5	5.9	298050	1	BX321861 Nitrosomo
C 45	236	5.8	335862	1	AP005089 Nitrosomo

ALIGNMENTS

RESULT 1


```

QY 561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
Db 1681 CTCTACCGCAACCGCTTGGCACTACATTTAGCCCAACCTTAACAGCAGCGCGCG 1740
QY 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAla 600
Db 1741 CCCAAATCCATCGAAGACGACAGCAATGAAGCTCGTGCCTACACCAATCCGCTGCG 1800
QY 601 AspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgGlyGly 620
Db 1801 GACTTCTACCGCGGAGGAGGGAATCTACTTCAACCGACCGCGCTACCGCATCGCG 1860
QY 621 ValSerGlyAspTyrValArgGlyArgLysLeuLysLeuProSerLeuProGlyArgGlu 640
Db 1861 GTTTCGCGCGACTATGTACGAGCGCGTCTGAAACCTTGGCCCTTACCGCGGAGGAA 1920
QY 641 AspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPro 660
Db 1921 GATGCTACGGACCGGCTCTTTCTATCGCGAGACGACCAAAACGCGCTCGCGTTCCG 1980
QY 661 AlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeu 680
Db 1981 GCTGCGCGCTCGGCTTCCACTGAAAGCTCGCTACCGACCGCATCGATGCCAATTG 2040
QY 681 AspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
Db 2041 GACTACTACCGGCTGTTTGGCCAAACAACTCGCCCGCTACGAAACGCGCACGCCCGGA 2100
QY 701 HisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAsn 720
Db 2101 CACCATATGCTCAACTCGCGCGCAACTTACCGCGCAATACGCTATGGCGAGTGGAA 2160
QY 721 TrpTyrValLysAlaAspAsnLeuLysGlnSerValTyrAlaHisSerSerPheLeu 740
Db 2161 TGGTACGTCAAGCGCAGCAACCTGCTCAACCAATCCGTTTACGCGCCACGACGCTTCTC 2220
QY 741 SerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2221 TCTGATACGCCCAATGGCGCGAGCTTTTACCGGTGCGGTAAACGTGAAGTTT 2274

RESULT 2
AX548017 2295 bp DNA linear PAT 26-NOV-2002
LOCUS Sequence 7 from Patent WO02060936.
DEFINITION AX548017
ACCESSION AX548017
VERSION AX548017.1 GI:25813126
KEYWORDS
SOURCE
ORGANISM
Neisseria gonorrhoeae
Neisseria gonorrhoeae
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1. Apicella, M.A., Edwards, J.L., Gibson, B.W., Scheffler, K. and Brown, E.
AUTHORS Vaccine and compositions for the prevention and treatment of
TITLE neisserial infections
JOURNAL UNIVERSITY OF IOWA RESEARCH FOUNDATION (US) ; University of
Patent: WO 02060936-A 7 08-AUG-2002;
California, Los Angeles (US) ; Apicella, Michael A. (US) ; Edwards,
Jennifer, L. (US) ; Gibson, Bradford W. (US) ; Scheffler, Karoline
(US)
FEATURES
Location/Qualifiers
1..2295
/organism="Neisseria gonorrhoeae"
/mol_type="genomic DNA"
/db_xref="taxon:485"
BASE COUNT 606 a 724 c 553 g 412 t
ORIGIN
Alignment Scores:
Pred. No.: 6.4e-229 Length: 2295
Score: 3928.00 Matches: 744
Percent Similarity: 98.68% Conservative: 5
Best Local Similarity: 98.02% Mismatches: 9

```

```

Query Match: 97.32% Indels: 2
DB: 6 Gaps: 0
US-09-936-377-2 (1-758) x AX548017 (1-2295)
QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db 19 ATGGCAACAACATACACTCAACCCATTTGTTTATCAATTCCTTTAATCAACACACCCCTC 78
QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
Db 79 CTTCTCCACAGCGATGGAACCTGACCAATCAGTGGGCTTGAAACCGGTGACGCTCGTGGC 138
QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db 139 AAAAGCGCTCGCGCGCCACTTCGGGGCTGCTGCACACTTCTACCGCTTCCGACAAAATC 198
QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db 199 ATCAGCGCGGACACCTTGGCAAAAAGCGTCACTTGGGTGATGCTTTAGACGCGTA 258
QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
Db 259 CCGGGCATTCATGCTCGCAATAGCGCGCGCGCATCGCTCCGCTTATTCGCGTCAA 318
QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
Db 319 ACAGGCGAGACGGATTAAGTGTGAACCATCAACGCGGCAACCGGCGACATGGCGGACTTC 378
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
Db 379 TCTCCAGACCATGCAATCATGTTGGACAGCGGCTTGTGCAACAGCTCGAATCCTGCGC 438
QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
Db 439 GGTTCGGTTTACGCTCTTGTACAGCTCGGCAATGTGCGGGGTGCTGCTCATGTTGCGCAT 498
QY 161 GlyValIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
Db 499 GGCATAATCCCGAAAAATGCTGAAACGGGCTATCGGCGCAACTCGATTCGATTCGTTTG 558
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
Db 559 AGCAGCGGCAATCTGGAATAACTCACTCGCGCGCATCAATATCGTTTGGSCAAAAC 618
QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
Db 619 TTTGTATTCACACGAGAGGCTGTACCGCAATCGGGGATTTACCGCGGTACCGGCTTAC 678
QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
Db 679 CGCAATCTGAACCGCTCGCGCGACAG-CCACGCGCGATTCGCAACCGGCGACATCGGCT 737
QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgAspGlnTy 260
Db 738 GTCTTGGTTGGGAAAAAGGCTTTTATCGCGCGCATACAGCGACCGCTCGCGACCAATA 797
QY 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleThrGlnLys 280
Db 798 TGTCTGCTGCTGCCCAACGACCAACGATGATTCGCGCGCATCATCTGCGCAAA 857
QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
Db 858 GAGTTTGTATTAAACAACGCTATTGTCAGCTTTTATCGCACCTTTCGCGAGACAT 917
QY 300 LAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
Db 918 CGATTACGACCAATCCCGGCTTGGCTGCGGCTTTTACGACGACGATGATGACACACCCCA 977
QY 320 sAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG 340
Db 978 TGCCCAACAGCGCAAAACCTTGGATGACCTCGCGCAACCAACGCTACGAACCTCGCGCGCA 1037

```

QY 340 uTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAspTyr 360
 DB 1038 ATGGAAGCAACCGTTCCCGGTTTGAAGCCCTCGCGGTACACCTGGAACCCGCAACGACTA 1097
 QY 360 rHisHisaspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 DB 1098 CGCCACGACGAAAGAGCGAGTGCAGTAGAAAACCTTTTAAACACCAACGCAAAA 1157
 QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGI 400
 DB 1158 CGCCCGTATCGAGTTGCGCCACCAACCCATAGCCGCTCTGAAAGGCGAGCTGGGGGTGCA 1217
 QY 400 nTrpLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProWe 420
 DB 1218 ATATTTGGACAAAAATCCAGTGTCTTATCTGCCCATCTCCGAGCGGTCAACACCCGAT 1277
 QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
 DB 1278 GCTGCTTGACATTAAGTGCACATTTACAGCTTTTTCGGTGTAGAACAGGCAACCTGGGA 1337
 QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 DB 1338 CAATCTCAGCTTGAGGCGGGGTAGCCGTGGAATAACAAAAAGGCTCCATCCGTACGA 1397
 QY 460 pLysAlaLeuIleAspArgGluAsnTyrTrpLysGlnProLeuProAspLeuGlyAlaHi 480
 DB 1398 CAAAGCATTAATGATCGGGAACAACTACTACAACCATCCCTGCGCGCACTCGCGCGCA 1457
 QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTrpPheThrProGlnHisIy 500
 DB 1458 CGCCCAACCGCCGCTCAATTCGCACTTTTCGGCAACTGGTATTTACGCCCAACACAAA 1517
 QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
 DB 1518 ACTCAGCTGACCGCTCCCATCAGGAACGCTCGCTCAACGCAAGAGCTGTACGCACA 1577
 QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
 DB 1578 CGCAACACAGCTCGCCACCAACACCTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG 1637
 QY 540 sSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAl 560
 DB 1638 TTCCAAATATCGAATCGCGCTGGGCTACGAGGCGACCGCTGCGCAATACAACTGGC 1697
 QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGI 580
 DB 1698 ACTCTACCGCAACCGCTTCGGCAACTACATTTACGCCCAACCTTTAAACGACGACGCGG 1757
 QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 DB 1758 CCCCAAATCATCGAAGACGACAGCGAATGAAGCTCGTGGCTACACCAATCCGGTGC 1817
 QY 600 aAspPheTyrGlyAlaGluGlyIleTyrPheLysProThrProArgTyrArgIleGI 620
 DB 1818 GGAATTTCTACGGCGGAGCGGAAATCTACTTTCAACCGCACACCGCGCTACCGCATCGG 1877
 QY 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGI 640
 DB 1878 CGTTTCGGCGGATATGTACAGCGCTCTGAAAAACCTTCCCTTCCCTACCGCGCAGGA 1937
 QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 DB 1938 AGACGCTACGGCAACCGCCCACTCATTTGCGCAAGCGCAACCAACCGCCCTCGCGTTCC 1997
 QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnIle 680
 DB 1998 GCGTCGCGGCTCGCGCTCCACTTGAAAGCTCGCTGACCGCGCATCGATGCGCAATTT 2057
 QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGI 700
 DB 2058 GGACTACTACCGGTGTTCGCCCAACAACTCGCCCGCTACGAAACGCGCAGCGCCGG 2117
 QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAs 720

DB 2118 ACACCATATGCTCAACTCGCGCAAACTACCGCGCAATACCGCTATGCGAGTGGAA 2177
 QY 720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPhele 740
 DB 2178 TTGGTAGCTCAAAAGCGGACAAACCTGTCTCAACCAATCGTTTACGCCACACGACGAGTCT 2237
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 DB 2238 CTCTGATACCGCGCAATGCGCGCAGCTTTTACCGCGCGCTGAACGTGAAGTTT 2292
 RESULT 3
 NMA422491/c 331801 bp DNA linear BCT 02-SEP-2002
 LOCUS Neisseria meningitidis serogroup A strain Z2491 complete genome;
 DEFINITION segment 4/7.
 ACCESSION AL162755 AL157959
 VERSION AL162755.2 GI:7379742
 KEYWORDS
 SOURCE Neisseria meningitidis Z2491
 ORGANISM Neisseria meningitidis Z2491
 Bacteria: Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 REFERENCE 1 (bases 1 to 331801)
 AUTHORS Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
 Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
 Davies,S.R., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
 Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
 Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
 Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
 Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis Z2491
 JOURNAL Nature 404 (6777), 502-506 (2000)
 MEDLINE 20222556
 PUBMED 10761919
 REFERENCE 2 (bases 1 to 331801)
 AUTHORS Parkhill,J.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
 COMMENT
 Notes:
 Details of N. meningitidis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
 FEATURES
 source
 1. 331801
 /organism="Neisseria meningitidis Z2491"
 /mol_type="genomic DNA"
 /strain="Z2491"
 /db_xref="taxon:122587"
 /note="serogroup: A"
 117. .1333
 /gene="NMA1040"
 /pseudo
 117. .1333
 /gene="NMA1040"
 /note="NMA1040"
 /note="possible type I restriction-modification
 system specificity protein, pseudogene, len: 1217 bp;
 N-terminus shows weak similarity to the C-terminal half of
 TR:Q50359 (EMBL:U25415), hsdS18, Mycoplasma pulmonis
 restriction-modification enzyme subunit S1B (336 aa),
 fasta scores; E(): 2.3e-06, 26.2% identity in 141 aa
 overlap and to many hypothetical restriction-modification
 subunits. Also similar to NMA1041, fasta scores; E():
 4.8e-10, 37.0% identity in 100 aa overlap. C-terminus
 similar to part of SW:U151 ECOLI (EMBL:X13145), hsdS,
 Escherichia coli type I restriction enzyme EcoR124II
 specificity protein (410 aa), fasta scores; E(): 1.9e-14,
 39.6% identity in 182 aa overlap. Similar to NMA1040, E():
 2.9e-09, 38.0% identity in 100 aa overlap. Contains pfam
 match to entry PF01420 Methylase_S, Type I restriction
 modification DNA specificity domain. Contains a G(8) tract

Query Match:	97.32%	Indels:	2
DB:	1	Gaps:	0
US-09-936-377-2 (1-758) x NMA422491 (1-331801)			
QY	1	MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu	20
DB	115608	ATGGCACAACCTACACTCAACCCCATGTTTATCAATTCCTTTTAAATCAACACACCCCTC	115549
QY	21	LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly	40
DB	115548	CTCTCCCAAGCGCATGGAACCTAGCAATCAGTGGCTTGGAAACGGTCAGCGTCGTGGC	115489
QY	41	LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle	60
DB	115488	AAAAGCCGTCGCGGCCACTTCGGGGCTGTGCACACTTCACCGCCCTCCGACAAAATC	115429
QY	61	IleSerGlyAspThrLeuArgGlnLysAlaValAlaLeuLeuGlyAspAlaLeuAspGlyVal	80
DB	115428	ATCAGCGCGCACCTTCGCACAAAAAGCCGTCAACTTGGGTGATGCTTTAGACGGCGTA	115369
QY	81	ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln	100
DB	115368	CGGGCAATTCATGCTCGCAATAACGGCGCGCGCATCCGCTCCCGTTATTTCGGGTCAA	115309
QY	101	ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe	120
DB	115308	ACAGGCAGACGGATTAAAGTGTGAACCATCAACGGCGCAACCGGGCGACATGGCGGACTTC	115249
QY	121	SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg	140
DB	115248	TCTCAGACCATCGAATCATGTGTGACAGCGCTTGTGCACACAGTCGAATCTCGCGC	115189
QY	141	GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp	160
DB	115188	GGTCCGGTTACGCTCTGTACAGCTCGGGCAATGTGGCGGGCTGTGTGATGTTCCGAT	115129
QY	161	GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgIleu	180
DB	115128	GGCAAAATCCCGGAAAAATGCTCGAAACCGCGTATCGGGCGAATCGGATTTGGGTTG	115069
QY	181	SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn	200
DB	115068	AGCAGCGCAATCTGGAAAACTCAGCTCCGGCGCATCAATATCGTTTGGGCAAAAAC	115009
QY	201	PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr	220
DB	115008	TTTGTATTGCACACGGAAGGGCTGTACCGCAATCGGGGATTAACCGGTACCGCGTTAC	114949
QY	221	ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla	240
DB	114948	CGCAATCTGAAACGGCTGCCGACAG - CCACCGCGATTTCGCAACGGGCGCATCGGGCT	114890
QY	241	ValLeuGlyTyrArgLysArgPheTyrArgArgThr - TyrSerAspArgArgAspGlnTyr	260
DB	114889	GTCTTGGGTGGCGAAAAAGGCTTTATCGGCGACACATACACGACCGTCCGACCAATA	114830
QY	260	rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys	280
DB	114829	TGGTCTGCTGCCACAGCCACGAATACGATGATTGCCACGCGCATCATCTGGCAAAA	114770
QY	280	sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa	300
DB	114769	GAGTTTGATTAAACAAACGCTATTTCGAGCTTTATCCGACCTGTGTGACGGAAGAAGCAT	114710
QY	300	IaspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis	320
DB	114709	CGATTACACAATCCGGGCTTGAGCTGGCGCTTTTCAACGACGACGATGATGCACGCCCA	114650
QY	320	sAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG1	340
DB	114649	TGCCCAACACGGCAAAACCTTGGATAGACTTGGCAACAAACGCTACGAACCTCCGGCGGA	114590
QY	340	uTrrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr	360
DB	114589	ATGGAAGCAACCGTTTCCCGGTTTTTGAAGCCCTGCGGTACACCTGAACCGCAACGACTA	114530
QY	360	rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs	380
DB	114529	CGCCACACGCAAAAAGCAGCGATGTCAGTAGAAACCTTTTAAACCAACCAACGCAAAA	114470
QY	380	nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValG1	400
DB	114469	CGCCCGTATCGATTGCGCCCAACCACTAGCCCGTCTGAAAGCGAGCTGGGGGTGCA	114410
QY	400	mTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe	420
DB	114409	ATATTGGGACAAAAAATCCAGTGTCTTATCTGCCACATCCGAAGCGGTCAACCAACCGAT	114350
QY	420	tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs	440
DB	114349	GCTGCTTGACAAATAAAGTGCAACATTACAGCTTTTTCGGTGTAGAACAGGCAAACTGGGA	114290
QY	440	pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs	460
DB	114289	CAACTTCAGCTTGAAGCGGGGTACGGTGGAAAAACAAAAAGGCTCCATCCGTACCA	114230
QY	460	physAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi	480
DB	114229	CAAGCAATTGATTGATCGGAAAACTACTCAACCATCCCTGCGCGACCTCGCGCGCA	114170
QY	480	sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLy	500
DB	114169	CGCCCAAAACCGCCGCTCATTCGCACCTTCGGGCAACTGGTATTTACGCCCAACACAAA	114110
QY	500	sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi	520
DB	114109	ACTCAGCTTGACCGCTCCCATCAGAACGCTGCGGTCAACGCAAGAGCTGTACGCACA	114050
QY	520	sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr	540
DB	114049	CGGCAAAACACGTGCGCACCAACACCTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG	113990
QY	540	sSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl	560
DB	113989	TTCCAAACAAATATCGAACTCGCGCTGGGTACGAGGCGACCGCTGGCAATACAACTGGC	113930
QY	560	aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgG1	580
DB	113929	ACTCTACCGCAACCGCTTCGGCAACTACATTACGCCCAACCTTTAACGACGAGCGCG	113870
QY	580	yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl	600
DB	113869	CCCCAAATCCATCGAAGACGACAGCGAAATGAAGCTCGGTGCGCTACAACTCCGGTGC	113810
QY	600	aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG1	620
DB	113809	GGACTTCTACGGCGCGGAAGGGAATCTACTTTAAACCGACACCGCGCTACCGCATCG	113750
QY	620	yValSerGlyAspTyrValArgGlyArgLysAsnLeuProSerLeuProGlyArgG1	640
DB	113749	CGTTTCCGGCGACTATGTACGAGCGCGTCTGAAAACTTCCTCCCTACCGCGCAGGA	113690
QY	640	uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr	660
DB	113689	AGACGCTTACGGCAACCGCCCACTCATTTGCCCAAGCCCAACCAACGCGCTCGGTTCC	113630
QY	660	oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe	680
DB	113629	GGCTGGCGCGCTCGCGTCCACCTTGAAAGCTTCGCTGACCGCGCATCGATGCCAATTT	113570
QY	680	uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG1	700
DB	113569	GGACTTACTACCGGTGTTCCGCCCAACAAACTCGCCCGCTACGAAACGCGCAGCGCCGG	113510
QY	700	yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrpAs	720

Db	113509	ACACATATGCTCAACCTCGCGCGCAAACTACCGCGCAATACCGCTATGGCGAGTGAA	113450
Qy	720	nTrrTyrValIysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe	740
Db	113449	TTGGTACGTCAAGACGCAACCTGCTCAACCAATCCGTTAGCGCCACAGAGCTTCCT	113390
Qy	740	uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValIleAsnValIysPhe	758
Db	113389	CTCTGATACGCGCAATGGGCGCAGCTTTACCGCGCGCGTGAACGTGAAGTTT	113335
RESULT 4			
AX300375			
LOCUS	AX300375	2277 bp	DNA linear PAT 30-NOV-2001
DEFINITION	Sequence 201 from Patent WO0185772.		
ACCESSION	AX300375		
VERSION	AX300375.1	GI:17381766	
KEYWORDS			
SOURCE	Neisseria meningitidis		
ORGANISM	Neisseria meningitidis		
REFERENCE	1		
AUTHORS	Tang, C.U.		
TITLE	Virulence genes, proteins, and their use		
JOURNAL	Patent: WO 0185772-A 201 15-NOV-2001;		
	Microscience Limited (GB)		
FEATURES	Location/Qualifiers		
source	1..2277		
	/organism="Neisseria meningitidis"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:487"		
CDS	1..2277		
	/note="unnamed protein product"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAD13141.1"		
	/db_xref="GI:17381767"		
	/translation="MAQTTLKPIVLISILLINTPLLAQHETEQSDVLETWSVVGKSRP		
	RATSGLLHTSASDKISGDTLRQKAVNLGDALDVGPIHASQYGGGASAPVIRGQTG		
	RRIVLNHGHTGDMADPSDPAHIMVDTALSQQVEILIRFPVTLLYSSGNAGLVADVAD		
	GKIPEKPNPENGVSSELGLRLSSGNLEKLTSGGINIGLGNFVLTHTGLYRKSGDYAVP		
	RYRNKLPLDSHADSGTSGISGLSWGKEGFTIGVAYSDRDQYGLPASHSEYDGDCHADI		
	IWQSLNKRVIQIYVPHLLTREDIDYNPGLSCGFHDDDDNAHATHSGRPWIDLRNKR		
	VELAEWKPQPPGEALREVLHNRDNDYHDEKAGDAVENFNNQTNARIEIURHQPIGR		
	LKGSWQVQLQOKSSALSATSEAVKQPMLLDNKVQHSYFFGVQEQANMDNFTLEGVRY		
	EKQKASIQYDKALIDRENYNHPDLPDGAHRQTARSFALSNGWYFTFOHKLSTASHQ		
	ERLPSTQELYAHGHKVAHTNTEPVGKHLNKRSSNNIELALGYEGDRWQYNLALYRNF		
	GNITYAOTLNDGRPKSTEDDSMKLVRYNKGADFYGAEGEIVFKPTPRYRIGVSGD		
	YVGRGLNPLSPGREDAYGNRPFIADDDONAPRPAARLGSPHLKASILTIDANLDY		
	YRVPAONKLARVETETPQHMLNLGANYRNRTRIGENWNYVKADNLLNQSVIAHSFL		
	SDTFQMGSRSTGGVNVKF"		
BASE COUNT	600 a	546 g	419 t
ORIGIN			
Alignment Scores:			
Pred. No.:	7, 28e-229	Length:	2277
Score:	3927.00	Matches:	743
Percent Similarity:	98.81%	Conservative:	7
Best Local Similarity:	97.89%	Mismatches:	8
Query Match:	97.30%	Indels:	2
DB:	6	Gaps:	0
US-09-936-377-2 (1-758) x AX300375 (1-2277)			
Qy	1	MetAlaGlnThrThrLeuIysProIleValLeuSerIleLeuLeuIleAsnThrProLeu	20
Db	1	ATGGCAACAACACTACACCAACCATCTGTTTATCAATCTTTTAAATCAACACACCCCTC	60
Qy	21	LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValValGly	40
Db	61	CTCGCCCAAGGCGATGAACCTGAGCAATCGGTGGATTGGAAACGGTCAGCGTCGTGGC	120

Qy	41	LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle	60
Db	121	AAAAGCCGTCGCGCGCACCTCGCGGCTGTGGCACTTCGACCGCTCCGACAAAATC	180
Qy	61	IleSerGlyAspThrLeuArgGlnIysAlaValIleAsnLeuGlyAspAlaLeuAspGlyVal	80
Db	181	ATCTCCGCGGATACCTTCGCGCAAAAGCCGTCAACTTGGGCGACGCTTACAGCGCGTA	240
Qy	81	ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln	100
Db	241	CGGCGCATCCACGCTTCGCAATACGCGCGCGCGCTCTGCTCCCGTCATTCGCGGTCAA	300
Qy	101	ThrGlyArgArgIleLysValIleuAsnHisIshisGlyGluThrGlyAspMetAlaAspPhe	120
Db	301	ACAGCGAGCGGATTAAGTAGTGTGAACCATCACGCGGAAACAGGCGATATGTCGGGATTT	360
Qy	121	SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg	140
Db	361	TCGCCCCGATCAGCCCATTTATGATACCGCTTGTGCGAACAGGTGCAATCTCTGGC	420
Qy	141	GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp	160
Db	421	GGCGCGGTTACGCTCTTGTACAGCTCGGCGCAATGTGGCGGGCTGTCGTCGATGTTGCCGAT	480
Qy	161	GlyIysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu	180
Db	481	GGCAAAATCCCGGAAAAATGCTGAAACGCGGTATCGGCGGAACTCGGATTGCGTTTG	540
Qy	181	SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn	200
Db	541	AGCAGCGCAATCTGGAAAAAATCAGTCCGCGCGCATCAATATCGGTTTGGGCAAAAC	600
Qy	201	PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr	220
Db	601	TTTGTATTGCACACGAGGCGTGTACCGCAAACTCGGCGGATTAACCGCTACCGCGTTAC	660
Qy	221	ArgAsnLeuLysArgLeuLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla	240
Db	661	CGCAATCTGAAACGCTGCCCGCAG-CCACGCGGATTCGCAAAACGGCGAGCATCGGCT	719
Qy	241	ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr	260
Db	720	GTCTTGGTTGGCAAAAGGTTTTATCGGCTGACGCTACGCGCGCGTCCGCGCAACAATA	779
Qy	260	rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrrGlnIy	280
Db	780	TGGTCTGCTCCCCACACGACGATACGATGATGCCACGCCGACATCACTCGGCAAAA	839
Qy	280	sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisIleuLeuThrGluGluAspVa	300
Db	840	GAGCTTGATTAAACACGCTATTACAGCTTTATCCGCACTTGTTCACCGAAGAAGACAT	899
Qy	300	LasPtyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis	320
Db	900	CGATTACGCAATTCGGGCTTGAGCTGCGGCTTCACGACGACGATAATGACACGCGACA	959
Qy	320	sAlaHisAsnGlyLysProIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGly	340
Db	960	CACCCACAGCGGCGACCGTGGATAGACCTCGCAACAAACGCTACGAACTCCGTCGCGA	1019
Qy	340	uTrrPlysGlnProPheProGlyPheGluAlaLeuArgValHisIleuAsnArgAsnAspTy	360
Db	1020	ATGGAAGCACCGTTCCCGGTTTTGAAGCCCTCGCGGTACACCTGAACCGCAACGACTA	1079
Qy	360	rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs	380
Db	1080	CGGCCACGACGAAAGACGCGCATGCGTCAAAATCTTTTAAACCAACCAACGCAAAA	1139
Qy	380	nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrrGlyValGly	400
Db	1140	CGCCCCGATCGAGTTGCGCACCAACCCATAGTGTGCTGGAAGGCGACGTCGGGCGTGCA	1199
Qy	400	nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe	420

```

|||||
1200 ATATTTACAAACAAAATCCAGTCTCTTATCGCATATCCGAAGCGGTAAACACCGAT 1259
|||||
420 tleuLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAs 440
|||||
1260 GCTGCTTGACAAAGATGCAACATTACAGCTTTTCGGTGTAGAACAGGCAACTGGGA 1319
|||||
440 pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
|||||
1320 CAACTTCAGCTTGAAGAGCGGTACGCGTGGAAACAAAACAAAAGCCTCCATTCACTACGA 1379
|||||
460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
|||||
1380 CAAAGCATTCATTGATCGGAAACACTACTCAACCCCTGCCACCTCGCGCGGCA 1439
|||||
480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLys 500
|||||
1440 CGCCAAACCGCGCTCATTCGCACTTCGGCACTCGTATTTACGCGCACACACAA 1499
|||||
500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
|||||
1500 ACTCAGCTTCAGCGCTCCCATCAGGAACGCGTCCGCTCAACGCAAGAGCTGTACGCACA 1559
|||||
520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
|||||
1560 CGGCAACACGTCGCCACCAACACCTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG 1619
|||||
540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
|||||
1620 TTCCAACAATATGAACTCGCTCGCTGGCTACGAAGCGGACCGCTGGCAATACATCTGGC 1679
|||||
560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGl 580
|||||
1680 ACTCTACCGCAACCGCTTCGCTAACTACATTATACGCCCAACCTTAAACGACGACGCGG 1739
|||||
580 YProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
|||||
1740 CCCCAAATCCATCGAAGACGACGAGCAATGAAGCTCGTGGCTACCAACCAATCCGCGCG 1799
|||||
600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGl 620
|||||
1800 CGACTTCTACGGCGGGAAGCGGAAATCTACTTCAAAACCGACACCGCGCTACCGCATCGG 1859
|||||
620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGl 640
|||||
1860 CGTTTCGGCGACTATGTACAGGCGCGTCTGAAACCACTCCCTTCCCTACCCGCGCAGAGA 1919
|||||
640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
|||||
1920 AGATGCTACGGCAACCGTCTTTCATCGCACAGGACGACCAAAATGCCCCCGCTGTTC 1979
|||||
660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
|||||
1980 GGCTGCGCGCTCGGCTTCACCTGAAAGCGCTCGCTGACCGACCGTATCGATGCCAATTT 2039
|||||
680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl 700
|||||
2040 GGACTACTACCGGTGTTCCGCCAAACAAACAACTCGCCCGCTACGAACGCGCACGCCCGG 2099
|||||
700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrpAs 720
|||||
2100 ACACCATATGCTCAACCTCGCGCAAACTACCGCGCAATACCGCTATGCGCAGTGGAA 2159
|||||
720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
|||||
2160 TTGGTACGTCAAAAGCGCACAACTGCTCAACCAATCGTTTACGCCCCACACGACGCTTCT 2219
|||||
740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
|||||
2220 CTCTGATACCGCGCAATGGCCGCGACCTTTACCGCGCGCGTGAACGTGAAGTTT 2274

```

RESULT 5
AE002448/c

LOCUS
DEFINITION
Neisseria meningitidis serogroup B strain MC58 linear BCT 25-MAY-2000 the complete genome.
ACCESSION
AE002448 AE002098
VERSION
AE002448.2 GI:7413446
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis MC58
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1 (bases 1 to 10700)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citterone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizzia, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R. and Venter, J.C.
Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58
Science 287 (5459), 1809-1815 (2000)
JOURNAL
MEDLINE
20175755
PUBMED
10710307
REFERENCE
2 (bases 1 to 10700)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citterone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizzia, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R. and Venter, J.C.
Direct Submission
Submitted (17-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
COMMENT
On Apr 4, 2000 this sequence version replaced gi:7226204.
FEATURES
source
1..10700
/organism="Neisseria meningitidis MC58"
/mol_type="genomic DNA"
/strain="MC58"
/db_xref="taxon:122586"
/note="serogroup: B"
complement(116..2392)
/gene="NMB0964"
complement(116..2392)
/gene="NMB0964"
/notes="similar to GP:2114321 percent identity: 43.79; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAF62323.1"
/db_xref="GI:7413447"
/translation="MAQTTLKPVLISILLINTPLLAQAHEATEQSDLETVSVVVKSRP
RATGLVHTSTASDKIISGDTLRQKAVNLGDALDGVGIHASQYGASAPVIRGQTG
RIKVLNHHGCTGDMADFPDHALMWDALSOQVEILRGVTLIYSSGNVAGLVVDVAD
GKIPEKPNHNGVSGEIGLSSNLEKLTSGGINIGLKNFVLHTGLYKRSQDYAVP
RYNLRKLPDSDHADSQTGSLGWGEKGFVIGVYDRDRDQYGLPAHSHEYDCHADI
IWOKSLINKRYLQYLPHELTEDIDYDNPGLSCGFHDNDNAHATHSRPIWIDRNRK
YILRAWKQPFQGFALRYHLNRNDYHDKAGDAVENFNQNTQARIELRHPTGR
LKGSMGVOLQOKSSALSALSAVKOPMLLDNKVCHYSPFGVQAWNDNTELEGGVRY
EQKASIQDKALIDRENYNPLPDLAGHRQTRASFALSGNWFPTPQHKLSLTASHQ
EKLPSTQELAHNGKVAITNFEVGNKLNKRSNNTELALGEGEDRWQYNLALYRNF
GYRITTAQTLNDGKPKSIDDSMKLVRYNQSGADFYGAEGEYFPTPRYICVSGD
YVYRGLKNLPSLPRFEDAYGNRPFFAQDDQNAIPVPAARLGFHLKSLATRIDANLDY
YRVFAONKLARYETRTPGHMLNLGANYRNTYRGWVYVKADNLNLQSVYHASFLL
SDTPQMGSRFTGTVNKKF"
complement(2450..2722)
/gene="NMB0965"

```

CDS      complement(2450..2722)
/ gene="NMB0965"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/ transi_table=11
/ product="hypothetical protein"
/ protein_id="AAF41370.1"
/ db_xref="GI:7226205"
/ translation="MRPREGAEPRVUTSTPFCHLFDPSGTCTFYALINFGYPAAEIN
GQQRNLQHQPRRTADTNGHRMPQPFHLKNFCLCPLSLKIL"
2814..3404
/ gene="NMB0966"
2814..3404
/ gene="NMB0966"
/ note="similar to GB:M32355 SP:P06193 GB:X02603 PID:I54228
PID:47816 percent identity: 85.95; identified by sequence
similarity; putative"
/ codon_start=1
/ transi_table=11
/ product="para-aminobenzoate synthase glutamine
amidotransferase component II"
/ protein_id="AAF41371.1"
/ db_xref="GI:7226206"
/ translation="MLLFDIDYDSFTYNIQVYFTELGOEVAVRENDITILEIEALNP
QYLIVGPGCSPKRAGISVAMRHFAGRLPIMGVCLGHQTLGEAFGGRIYRAKTLMEG
KVSFVSHSGKMGKGLNPVTCRYHSLVIDRNTMPECLVATWTDGEIMGVRHKEY
AVSGVQHPHALTEHGHDMNNFLIEFQNKFKQKI"
3468..4526
/ gene="NMB0967"
3468..4526
/ gene="NMB0967"
/ note="similar to GB:M33814 SP:P20574 PID:I51234 percent
identity: 68.05; identified by sequence similarity;
putative"
/ codon_start=1
/ transi_table=11
/ product="anthranilate phosphoribosyltransferase"
/ protein_id="AAF41372.1"
/ db_xref="GI:7226207"
/ translation="MTPQQAIELRLISNNELFYDEMDDLMRQMRGQVLPQIAAILT
GLRKIVTSEITAAAMVREFATKPLENAEGLVDIVGTGDCGAKTFNISTTSMFVA
AAAGAKVAKHGSSVSSGAADVVEQGANLNLTPQVAQISROTIGTGFMEAPNHS
AMRHVAPVRSILGFRSIFNLGLPTNPAGPNQLLGVPHTDLGILSRVLQQLGSKHV
LVVCGEGLEILTIGTRVABLKDGKISYDIRPEDFGIETRNLDIEIKVANTQSSL
LKQNEVLGREGAARDIVLENTAAALYAGNVAASLSDGISAAREAIDSGRAKSKREF
VGFQPOQRCHFLGKMEIG"
4534..4623
/ gene="NMB0968"
4534..4623
/ gene="NMB0968"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein"
/ protein_id="AAF41373.1"
/ db_xref="GI:7226208"
/ translations="MIMELWQNISTRRELTKFLYKNLGLY"
complement(4678..4851)
/ gene="NMB0969"
complement(4678..4851)
/ gene="NMB0969"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein"
/ protein_id="AAF41374.1"
/ db_xref="GI:7226209"
/ translation="MLPFIIRKKVKGIVTDFRSYDVLDISEFSLRKFNIGPIKEH
LGLHLKCKQWHFK"
4995..6419

```

```

/ gene="NMB0970"
/ note="This region contains an authentic frame shift and
is not the result of a sequencing artifact; conserved
hypothetical protein, authentic frameshift; identified by
Glimmer2; putative; conserved hypothetical protein,
authentic frameshift"
6382..6672
/ gene="NMB0971"
6382..6672
/ gene="NMB0971"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein"
/ protein_id="AAF41375.1"
/ db_xref="GI:7226210"
/ translation="MNILOGHIVGVKKINGOLEGKSFDCCLIVATPLDSSQGNALG
SSTIEDFGGSANFEQFRNAQFFIEANLNVEIVTIGTKQLKLVGFQLVKKG"
6677..6874
/ gene="NMB0972"
6677..6874
/ gene="NMB0972"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein"
/ protein_id="AAF41376.1"
/ db_xref="GI:7226211"
/ translation="MQKYVYVQSVSTGDFLYLSPETGDIGHTKLITNADYDFEEAL
NAGLEEIGNQYEFVVFGLND"
6882..7166
/ gene="NMB0973"
6882..7166
/ gene="NMB0973"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein"
/ protein_id="AAF41377.1"
/ db_xref="GI:7226212"
/ translation="MFGRLKNAPSIATAKHLKENIMKFINTCRKYGAKLAVVTAAPL
ALAAHANATIPDTAKNALEAAKADGMEAGNIVVIGFAALFVFSIVKRVMK"
7173..7451
/ gene="NMB0974"
7173..7451
/ gene="NMB0974"
/ note="hypothetical protein; identified by Glimmer2;
putative"
/ codon_start=1
/ transi_table=11
/ product="hypothetical protein"

Alignment Scores:
Pred. No.:      5,19e-228      Length:      10700
Score:          3927.00      Matches:      743
Percent Similarity: 98.81%      Conservative: 7
Best Local Similarity: 97.89%      Mismatches: 8
Query Match:      97.30%      Indels:      2
DB:              1              Gaps:      0

US-09-936-377-2 (1-758) x AE002448 (1-10700)

Qy      1 MetAlaGlnThrThrLeuTySProIleValLeuSerIleLeuLeuIleAseThrProLeu 20
Db      2392 ATGGCAAACTACATCAAAACCCATTGTTTATCAATCTTTTAAATCAACACACCCCTC 2333
Qy      21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
Db      2332 CTGCCCAAGCGCATGAACTGAGCACTCGGTGATTGGAAACGTCAGCGTGTGGC 2273

```

QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 DB 2272 AAAAGCGTCGCGCGCCACGTCGCGGCTGTTGCACACTTCGACCGCTCCGCAAAATC 2213
 QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 DB 2212 ATCTCGGCGATACCTTGGCCAAAAGCGCTCACTTGGGCGACGCTTTAGACGGGTA 2153
 QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 DB 2152 CGGGGATCCACGCTTCGCAATACGGCGCGCGCTGCTCCCGTCATTCGCGGTCAA 2093
 QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 DB 2092 ACAGGAGCGGGAATTAAGGTGTTGAACCATCACGGGGAACAGCGCATATGCGGGATTTT 2033
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 DB 2032 TCGCCCGATCACGCCATTATGTTAGATACCGCTTGTGCAACAGGTTCGAATTCCTGCGC 1973
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 DB 1972 GGGCCGGTTACGCTCTGTACAGCTCGGCAATGTGGCGGGCTGGTTCGATGTGCGCAT 1913
 QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 DB 1912 GGCAAAATCCCCGAAAAAATGCCGTGAACCGGCTATCGGGCGAATCGGAATTCGCGTTG 1853
 QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 DB 1852 AGCAGCGCAATCTGGAAAAAATCACGTCGCGGCGCATCAATATCGGTTTGGCAAAAC 1793
 QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
 DB 1792 TTTGTATTCACACGGAAGGGCTGTATCCGCAAAATCGGGGATTCACGCGGTACGCGTTAC 1733
 QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 DB 1732 CGCAATCTGNAACGCTGCCCGACAG-CCACGCGGATTCCGAAACGGGCGAGCATCGGGCT 1674
 QY 241 ValLeuGlyTrpArgLysArgPheTyrArgThr-TyrSerAspArgAspGlnTyr 260
 DB 1673 GTCTTGGTTGGCGAAAAAGGTTTATCGGCGTAGCGTACAGCGACCGTCGCGACCAATA 1614
 QY 260 rGlyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAspIleIleTrpGlnLys 280
 DB 1613 TGTCTCGCTGCCCAACAGCCAAATACGATGATTCGACGCGGACATCATCTCGGCAAAA 1554
 QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
 DB 1553 GAGCTTGATTACAAACGCTATTATCAGCTTTATCCGACCTGTTGACCGAAGACAT 1494
 QY 300 lAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHi 320
 DB 1493 CGATTACGAAATCCGGGCTTGAGCTGGCGCTTCCACGACGACGATAATGACACGACACA 1434
 QY 320 sAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG 340
 DB 1433 CACCCACGGCGACCGTGGATAGACCTTGGCCAAACACGCTACGAACTCCGTGCCGA 1374
 QY 340 uTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360
 DB 1373 ATGGAAGCAACCGTTCCTCCCGGTTTGAAGCCCTGCGGTACACTGAAACCGACACACTA 1314
 QY 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 DB 1313 CCGCCACGACGAAAAAGCGCATGCGTTCGAAAACTTTTAAACAAACGCAAAA 1254
 QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpClyValGI 400
 DB 1253 CGCCCCGATCGAGTTGCGGCCACCAACCCATGATGCTGTGAAGGAGCGTGGGGCGTGCA 1194
 QY 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProWe 420

RESULT 6
 AX044031/c

DB 1193 ATATTTTACAACAAAAATCCAGTGTCTTTATCTGCATATCCGAAGCGGTTAAACAACCGAT 1134
 QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
 DB 1133 GCTGCTTGACAAAGTGCACATTTACAGCTTTTTTCGGTGTAGAACAGAGCAAACTGGGA 1074
 QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 DB 1073 CAACTTCACGCTTGAAGGAGCGGTACGCTGGGAAAAACAAAAAGCCTCCATTTCAGTACGA 1014
 QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
 DB 1013 CAAAGCATTTGATTGATCGGAAAAACTACTACAACCCACCCCTCCCGACCTCCGCGCGCA 954
 QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLy 500
 DB 953 CGCGCAAAACCGCCCGCTCAITTCGCACTTTTCGGGCAACTGGTATTTTACGCCACCAACAA 894
 QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
 DB 893 ACTCAGCTGACCGCTCCCATCAGGAACGCTTCGCTCAACGCAAGAGCTGTACGCACA 834
 QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
 DB 833 CGCAAAACAGCTCGCCCAACACCTTTGAAGTCGCGCAACAAACACCTCAACAAAGAGCG 774
 QY 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
 DB 773 TTCCACAATATCGAACTTCGCTGGGCTACGAAGGCGCCCTGGCAATACAAATCTGGC 714
 QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGI 580
 DB 713 ACTCTACCGCAACCGCTTCGGTAACTACATTTACGCCCAAAACCTTAAACGACGACGCGG 654
 QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 DB 653 CCCCAATTCATCGAAGACGACAGCAAAATGAAGTCGTCGCTTACAAACCAATCCCGCGC 594
 QY 600 aAspPheTyrGlyValaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGI 620
 DB 593 CGACTTCTACCGCGCGGAGGCGAATCTACTTCAACCGACACCGCGCTACCGCATCGG 534
 QY 620 yValSerGlyAspTyrValArgGlyArgLysAsnLeuProSerLeuProGlyArgGI 640
 DB 533 CGTTTCCGCGCACTATGTACGAGCGCTCTGAAAAACCTGCCTTCCCTACCCGCGCAGAGA 474
 QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 DB 473 AGATGCTTACGGCAACCGCTCTTCATCGCACAGGACGACCAAAAATGCCCCCGTGTTC 414
 QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
 DB 413 GGCTGCGCGCTCGGCTTCCACTGAAAGCCTCGCTGACCGACCGTATCGATGCCAATTT 354
 QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGI 700
 DB 353 GGACTACTACCGCGTTCGCCCAAAACAACTCGCCCGCTACGAAACGCGACGCGCGG 294
 QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyuTrpAs 720
 DB 293 AACCCATATGCTCAACTCGGCGCAAACTACCGCCCAATACGCGCTATGCGCGAGTGGAA 234
 QY 720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 DB 233 TTGTAGGTCAAGCGCAACCTGCTCAACCAATCCGTTTACGCCCAACGACGAGCTTTCT 174
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 DB 173 CTCTGATACGCGCAAAATGGGCGCAGCTTTTACCGGCGCGTGAACGTCGAAGTTT 119

LOCUS AX044031 349980 bp DNA linear PAT 24-NOV-2000
 DEFINITION Sequence 110 from Patent WO0066791.
 ACCESSION AX044031
 VERSION AX044031.1 GI:111342915
 KEYWORDS
 SOURCE
 ORGANISM
 Neisseria meningitidis
 Neisseria meningitidis
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 1
 PIZZA, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C.,
 Maignani, V., Galeotti, C., Mora, A. M., Ratti, G., Scariselli, M.,
 Scarlato, V., Rappuoli, R., Frazer, C. M. and Grandi, G.
 Neisseria genomic sequences and methods of their use
 Patent: WO 0066791-A 110 09-NOV-2000;
 CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
 FEATURES
 source
 1. 349980
 /organism="Neisseria meningitidis"
 /mol_type="genomic DNA"
 /db_xref="taxon:487"
 /notes="sequence too long, cut in 8 pieces.--seq 1: 1 to
 349980 349980 bases--seq 108: 300001 to 649980 349980
 bases--seq 109: 600001 to 949980 349980 bases--seq 110:
 900001 to 1249980 349980 bases--seq 111: 1200001 to 1549980
 349980 bases--seq 112: 1500001 to 1849980 349980 bases--seq
 113: 1800001 to 2149980 349980 bases--seq 114: 2100001 to
 2272325 172325 bases"
 BASE COUNT 81351 a 86755 c 95584 g 86290 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.34e-226 Length: 349980
 Score: 3927.00 Matches: 743
 Percent Similarity: 98.81% Conservative: 7
 Best Local Similarity: 97.89% Mismatches: 8
 Query Match: 97.30% Indels: 2
 DB: 6 Gaps: 0
 US-09-936-377-2 (1-758) x AX044031 (1-349980)
 Qy 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuAsnThrProLeu 20
 Db 85503 ATGCACAAACTACACTCAAAACCATGTTTATCATCAATCTTTATCAACACACCCCTC 85444
 Qy 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
 Db 85443 CTCGCCCAACGCATGAAATGTAGCAATTCGGTCGATTTGGAAACGGTCAGCGTCGTCGC 85384
 Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 85383 AAAAGCGTCCGCGGCCAGCTCGGGCGTGTTCACACTTCGACCGCCCTCCGACAAATC 85324
 Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 85323 ATCTCCGGCGATACCTTCGCCAAAAGCGCTCAACTTCGGCGCGCTTTAGACGGCTA 85264
 Qy 81 ProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 85263 CCGGGCATCCACGCTTCGCAATACGGCGCGCGCGCTCTGCTCCGCTCATTCGCGTCAA 85204
 Qy 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAsp 120
 Db 85203 ACAGGCAGGCGGATTAAGTGTGAACCATCACGGCGGAACAGGCGGATATGGCGGATTT 85144
 Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 Db 85143 TCGCCCGATCACGCCATTATGGTAGATACCGCTTGTTCGCAACAGTCAAAATCTCTGGC 85084
 Qy 141 GlyProValThrLeuLeuThrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db 85083 GGSCCGGTACGCTCTGTACAGCTCGGCAATGTGGCGGGCTGTGTCGATCTGCCAT 85024

83944	CGGCAAAACAGTCGCGCAACCAACACTTTGAAGTCGCGCAACAACACCTCAACAAGAGCG	83885	
QY	540	gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl	560
Db	83884	TTCCAAACAATATCGAATCGCGCTCGGCTACCAAGCGACCGCTGCGCAATACAAATCTGGC	83825
QY	560	aleuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgG1	580
Db	83824	ACTTACCGCAACCGCTTCGGTAACTACATTTACGCCCAAACTTAAACGACGAGCGCGG	83765
QY	580	yProLysSerIleGluAspAspSerGluMetLysLeuValaTgTyrAsnGlnSerGlyAl	600
Db	83764	CCCCAAATCCATCGAAGACGACAGCAAAATGAAGCTCGTGCCTACAAACCATCCGGCGC	83705
QY	600	aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG1	620
Db	83704	CGACTTCTACCGCGCGGAGGCGAAATCTACTTCAAACGACACCGCGCTACCGCATCGG	83645
QY	620	yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgG1	640
Db	83644	CGTTTCCGGCGACTATGTACGAGGCGCTCTGAAAACCTGCCCTTCCCTACCCGCGCAGAG	83585
QY	640	uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr	660
Db	83584	AGATGCCTACGCGCAACCGTCCTTCATCGCAGCAGCAACAAATGCCCCCGTGTTC	83525
QY	660	oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe	680
Db	83524	GCTCGCGCGCTCGGCTTCCACTGAAAGCCTCGCTGACCGACCGATGATGCGCAATTT	83465
QY	680	uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG1	700
Db	83464	GGACTACTACCGCGTTCGCGCCAAACAACTCGCCCGCTACGAAACCGCGCGCGCG	83405
QY	700	yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyLTrpAs	720
Db	83404	ACACCATATGCTCAACCTCGCGCAAACTACCGCGCAATACCGCGCTATGCGGAGTGGAA	83345
QY	720	nTrpTyrValLysAlaAspAsnLeuLysGlnSerValTyrAlaHisSerSerPheLe	740
Db	83344	TTGGTACGTCAAGCGCGCAACCTGCTCAACCAATCCGTTTACGCCCAACGAGCTTTCT	83285
QY	740	uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValaAsnValLysPhe	758
Db	83284	CTCTGATACGCGCAAAATGGCGCGCAGCTTTACCGCGCGGAGCGTGAAGTGT	83230
RESULT 7			
AY028475			
LOCUS	4931 bp	DNA	linear BCT 29-JAN-2002
DEFINITION	Mannheimia haemolytica ribonuclease PH-like protein and iron-regulated outer membrane protein (irp) genes, complete cds.		
ACCESSION	AY028475		
VERSION	AY028475.1 GI:13591378		
KEYWORDS	Mannheimia haemolytica		
SOURCE	Mannheimia haemolytica		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Mannheimia.		
REFERENCE	1 (bases 1 to 4931)		
AUTHORS	Graham,M.R. and Lo,R.Y.		
TITLE	A putative iron-regulated TonB-dependent receptor of Mannheimia (Pasteurella) haemolytica A1: possible mechanism for phase variation		
JOURNAL	Vet. Microbiol. 84 (1-2), 53-67 (2002)		
MEDLINE	21589082		
PUBMED	11731159		
REFERENCE	2 (bases 1 to 4931)		
AUTHORS	Lo,R.Y. and Graham,M.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-MAR-2001) Microbiology, University of Guelph, Guelph, Ontario N1G 2W1, Canada		
FEATURES	Location/Qualifiers		

source	1..4931		
	/organism="Mannheimia haemolytica"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:75985"		
CDS	1063..1779		
	/codon_start=1		
	/evidence=not experimental		
	/transl_table=11		
	/product="ribonuclease PH-like protein"		
	/protein_id="AAK29744.1"		
	/db_xref="GI:13591380"		
	/translation="WRPNRRANNOVREIKITENYTRYAEGSVLIEFGETKVLGNATIVE		
	ESVPRFLGQQQGWTAIYGMPLPRAHSTQREAAKGGQGGRTMEIQRLIARSLRAVV		
	DUKALGERTVTVDDCVIQADGGTRTASTIGACVLYDANNKLIVANGTILTKTPMKGLVA		
	AISGVIVAGEAVCDLEYLEDNSAETDMNVVMVQDGLRVEVQGTAEGETPFSHOELLTL		
	DLAKGQISQOLVEAQRKALAE"		
gene	2003..4303		
	/gene="irp"		
CDS	2003..4303		
	/gene="irp"		
	/function="involved in iron acquisition"		
	/note="TonB-dependent"		
	/codon_start=1		
	/transl_table=11		
	/product="iron-regulated outer membrane protein"		
	/protein_id="AAK29743.1"		
	/db_xref="GI:13591379"		
	/translation="MLKKNYLTYSILLSAISGVYANEISLETITVDGNTPTSGKLLG		
	GELNSESVDENLKKQSGITLGNALSGELGTHSQFGGASTPIRQGESKRKILQ		
	NNGNLDMSGPHAVTVALLAKRIEILRGPTLLYSAGNTAGVINVDNKLPTAI		
	PEKVGEGFVGVRFGSASKEERTLYAGSTFALGNHLALRVQGMNKASEYIYAPHTIEGK		
	PYHRVPSDVQSQGTGTSLSWIGERHGLGIATYDRDKYGLIHTHKDHYDITISIRQ		
	AVMFAGYLRTFPLAEEGDIDYNNPGLRLHHTIPGGSHYQDTHRGKPMIDHMSK		
	RYIDGSLQNFPGFEEAKISANYVDYHDEKQGRVENYFNKGNKRLFERLHVHKEWK		
	GLKGAIGVQYTNQSTLSALASRAAKVFNKPLNNPKTKLMSLFAIRNLGDPFTF		
	ELSGRAERQKIAMDYDKLIDRWLGFTMPNLDHPKDKGYSYFATHWYFAPNKLIT		
	LNAAHQERLPNAOELYAHGKIALNAPEAGNKLKRSNQIELSLAYVGDKWDYKLN		
	LYHTRYGNYIYPLINDNRGPKSFTEYNLKNRYIQGEARFSGAEGISGLYLTFFNRY		
	LAVGDYVRGLVNLNPMNSYNIWIGVDKWSQDIPRIPPLRGARFNADPNL		
	NWSGMEKLYRVFAQKKVYQGVFTVHHQVNLGVTSYHNFQTEYQVFLKVDNLLNQK		
	MYQASLYLPHIPQMGKNAMLGWNTSF"		
BASE COUNT	1519 a	1022 c	957 g 1433 t
ORIGIN			
Alignment Scores:			
Pred. No.:	4.02e-95	Length:	4931
Score:	1725.50	Matches:	346
Percent Similarity:	60.61%	Conservative:	128
Best Local Similarity:	44.25%	Mismatches:	274
Query Match:	42.75%	Indels:	35
DB:	1	Gaps:	8
US-09-936-377-2 (1-758) x AY028475 (1-4931)			
QY	2	AlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeu	21
DB	1985	TCTAGAGAGACTCTAAACGTGTAAAGAAAAATATCTCAGTGATCAATTCCTCTGCC	2044
QY	22	AlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLys	41
DB	2045	ATTAGTGAGTAGGCTACGCTAAATGAAATCTCTTTAGAGACTATTACTGTTGATCGAAAT	2104
QY	42	SerArgProArgAlaThrSerGlyLeuLeuHisThrSer---ThrAlaSerAspLysIle	60
DB	2105	ACG---CCTTCAACCAAGGAAATATCTAGTGTGTGAATAAATAGCAATGAATCTGTT	2161
QY	61	IleSerGlyAspThrLeuArgGlnLysAlaValaAsnLeuGlyAspAlaLeuAspGlyVal	80
DB	2162	GTTGATGAAAAAATTAAACAAGGTAGTATTACATTAGGTAAACGCTTTATCTGTTGAA	2221
QY	81	ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln	100
DB	2222	TTAGGCATTTCATTCAGCCAGTTCGGAGCGGTGCAAGTACCCTCAATATTTCGTGGGCAA	2281


```

VERSION      AX052567.1  GI:12226784
KEYWORDS
SOURCE
ORGANISM      Moraxella catarrhalis
               Moraxella catarrhalis
               Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
               Moraxellaceae; Moraxella.
REFERENCE
AUTHORS      Thonhard, J. S.
TITLE        Novel compounds from moraxella catarrhalis
JOURNAL      Patent: WO 0071724-A 9 30-NOV-2000;
              SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES
source
1..2457
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"
BASE COUNT   744 a 550 c 554 g 599 t
ORIGIN
Alignment Scores:
Pred. No.:    5,35e-87      Length:      2457
Score:        1585.00      Matches:    338
Percent Similarity: 58.92%   Conservative: 141
Best Local Similarity: 41.57% Mismatches:    272
Query Match:  39.27%      Indels:      63
DB:           6           Gaps:       13

US-09-936-377-2 (1-758) x AX052567 (1-2457)
QY      7  LysProIleValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db      22 AAACCGCTTCCCTGTCGCAATATGGACATTTTCATGCAATGCGGAGGCGGAAT 81
QY      26 ---GluThrGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
Db      82 TTAAGGATAAGCAACCGTCATTTAGATGGCGTTTCGATCACCTCTTTAGTGACCAA 141
QY      45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspIleIleSerGlyAsp 64
Db      142 AATACAGAGTTTGGCGTTAATCATCTCAAAACACAGTCAGTGGCATCAAGTTTCAAAAGAG 201
QY      65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
Db      202 CAACTACACACAGGACCAACCCCTAGGCGATGCTTGGCAGTGAGCTTGGCGTTCAT 261
QY      85 AlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
Db      262 TCTAACCAATTTGGGGCGGTGCTCAGCCCCCATCATTCGTGGCAGGAGGTAAACGC 321
QY      105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
Db      322 CTGAATCTCTCAAAACCGTTGAGAGTTGTGGACATGTCGGGTTGTGCGCAGACCAT 381
QY      125 AlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArgGlyProValThr 144
Db      382 GCCATAGCGTGGACACCACTGGCAAAACAGGTGGAGATTGTGGAGGCTCTGGTGC 441
QY      145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
Db      442 TTGTGTAGCCCTCTGGCACTCAGCAGGCGTGGTCAATGCTGTTGATGACAAAATACCC 501
QY      165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgIleuArgSerGlyAsn 184
Db      502 AGCAATITGCCACGCAAA---TTACAGGTGATGTGACGGTGGCTCTTAGCATGGCCAAC 558
QY      185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
Db      559 CGTGAATAATTAATCACCGCAGTGGCGAAGCCCACTGGGAGAGCATGTGGCAGTGGCGT 618
QY      205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
Db      619 GTTGCAGGGGTGTCCAAACAGGACGACGACTATAAAACGCCGCTTTGACCGCCCATGTC 678

```

```

221 QY -----ArgAsnLeu 223
679 Db TTTTACAAAAACATGAAGATGATAACACTCAGCAGCAATTCATCTATAAGACACCTTA 738
224 QY LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
739 Db AAGCATCTGCCACAGCCCATGCCAATCAAAACGAGGAACGCT- TGGCGTGTGCATGGT 797
244 QY TTPArgLysArg-PheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuPr 263
798 Db TGGCAATCAAGGCTTTTGGGGCATCGGTGAGCTTACGCCGAGACAAATATGGCCTGCC 857
263 QY oAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuI 283
858 Db CAACCATTCACATGAATATAGCAATGTAGCGTGCATGGCATTTCTCAGTCCGCTTACA 917
283 QY eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAs 303
918 Db ATACAAGCCCATATTTGGCTTTGTATCTTTTGTGGAATAATGATGACTTAGAGTTGA 977
303 QY pAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHi 318
978 Db CAATGACGGCTTGAATGCCATACATACATGATGACACGACGACGACCATGCTCA 1037
318 QY sAlaHisAlaHis-----AsnGlyLysProTyrIleAspLeuArgAsnLys 333
1038 Db TGACCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1097
333 QY sArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgVa 353
1098 Db GCGTTATGATGTCAAGGCGCAATCAATGCGCGTGTGTGGCATTTGATATAATCCGAGC 1157
353 QY lHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPh 373
1158 Db CACCATGGGCAAAGTGGATTATCATCATGATGAGATAGATGGGGTGGAGACCACTT 1217
373 QY ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle----- 390
1218 Db TTTTGTATAATCAAGCCCAATGTGTGGCGTCTCGAAGCCTCAATACCCCATTCATACGCC 1277
391 QY ---GlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
1278 Db GATGGGCAAGTTTAGCGGGGTGTTTGGGGTAGGTTATCTCACCTCAAAAACAGCGGACT 1337
409 QY u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
1338 Db TGTGCCACCTCGTTATGACATGCGCAATAAACAAGACACCCCAACATCTTGCACATAA 1397
424 QY nLysValGlnHisTyrSerPheGlyValGluGlnAlaAsn---TrpAspAsnPheTh 443
1398 Db TAAACCAAAACACAGGCGAGTGTGTTTGGTTTGAAGAATAACAACCCCAATGACAAGCTGAC 1457
443 QY rLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLe 463
1458 Db CGTTGACGCGCGCTCGCATTTGAGAACAACCAATCACCATTGATGATATGATAAGACGC 1517
463 QY uIle-----AspArgGluAsnTyrTyrLy 471
1518 Db CATTTATCAGAGCTTAACCTTAGGCTTAGCAACCGCTCATGAACACGACATACGCTTAA 1577
471 QY sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
1578 Db ACGATTGCTGGAAGCGGTACTTTTAAACCCCAAAACAAACCGCAGCTCTTATGCTGT 1637
489 QY uSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl 509
1638 Db TGGGACGCAATTTACAATTAACGCCCAACATAAATATCGCTGAATCTGCTCGCATCAAGA 1697
509 QY uArgLeuProSerThrGlnLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
1698 Db ACGCTTGCCAAATGTCTCAGGAATTTGATGCTCAGCGCATGACCTTGCCACCACTCGTT 1757
529 QY eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549

```

```
Db      1758  TGAATTTGGCAACCGCTTTTAAACAAAGAAAATCCACACATTTGATTTGGGCTGAC 1817
Qy      549  yTyrGluGlyAspA-gTTrpGlnTyrAsnLeuAlaLeuTyrA-gAsnA-rPheGlyAsnTy 569
Db      1818  ATTTCAAGGTGATAAAATGGGATTCATCGTCTGGGGCTATCATATTGATTTTGATACTA 1877
Qy      569  rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerS-IleGluAspSe 588
Db      1878  TGTGTGTTTACAAACATTTCCGAGTATTAAGCAAGGTTTGGCGTCATCGTCATATAA 1937
Qy      588  rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyG 608
Db      1938  AGACITAAACCGCAGCGCTATGAACAGACGCGGGAATTTTATGGATTGATGCTCAA 1997
Qy      608  uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgI 628
Db      1998  CATCGGTTATACAGATTAATGATGATATATCATGTCGCGGTTATTGGTGATATATTC 2057
Qy      628  yArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnA-rGProPh 648
Db      2058  CAACTCTACCAATTTGCCCTGACAAAAGGCGAGAACCGATGCGTATGGCAACGCTCT 2117
Qy      648  eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668
Db      2118  CATCAACAGCGCAGACAGTCATAGCCCAAGACTGCCACCAAAACGCTTGGCATGAATT 2177
Qy      668  uLysAlaSerLeuThrAspA-rG-IleAspAlaSerLeuAspTyrTyrA-rGValPheAlaG 688
Db      2178  AACCGCCCATGTTAATGCAAAATGGTCAGGGTTTTTGGAAATATCGCATACCTTTAA 2237
Qy      688  nAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAl 708
Db      2238  AGATAAAATGGGAATTTTGACCGCCCAACCCAGCTCATTAAGTGGTGAATTTGGGCT 2297
Qy      708  aSerTyrArgA-rGAsnThrArgTyr-----GlyGluTrpAsnTrpTyrValLysAla 726
Db      2298  TAACTATCAGCACAAGCAAGCATCAAGCAGGCTCGGTTTCAGGTATTTTTAAATGCTAA 2357
Qy      726  pAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMe 746
Db      2358  CAATCTATTAAACGATAAAAGTCTTGCTCATGAGACATTTTCCCGACATGCCCAAA 2417
Qy      746  tGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db      2418  GGGCGAAACTTATCTCGGGCAAAATTCAAATTT 2454

RESULT 9
AX394914
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Moraxella catarrhalis
Moraxella catarrhalis
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Moraxellaceae; Moraxella.
1
REFERENCE
AUTHORS
Loosmore,S., Wang,J., Bradley,B., Ochs,M. and Yang,Y.P.
TITLE
Moraxella polypeptides and corresponding dna fragments and uses
thereof
JOURNAL
Patent: WO 0218595-A 51 07-MAR-2002;
Aventis Pasteur Limited (CA)
FEATURES
source
Location/Qualifiers
1..2646
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"
101..2545
/note="unnamed protein product"
/codon_start=1
/transl_table=11
CDS
```

```
/protein_id="CAD32083.1"
/db_xref="GI:21065988"
/translation="MIKKPLACAILATTFSPMLAEANLKDPTVILDGVSIITSLADQN
TEFGVNSKTVSGITVSKREQLQOARTLLDALGELGVHSHNPFGGASAPIIRGQEGK
RLKILQNGSEVVVMSGLSPDHAIAVDVTLAKQVEIVRSGALLYASNGAGVNVVDD
KIPSLPKLQGDVTPRLSSANREKLIITASAEPLGEHVAHVASGKAQADVKYTPRF
DRHFNKKHEDDNTQPEFYKDTLKLHLPDSHAKSNAGTLGVSWVGNQGLGASVLSRR
DKYGLPNHSHEYERESVHGISQSALQYKPYLRLYPLMENDDLLEFNNAGLECHTHDDH
DEHDHADDHDEHDEHDKPWIDLMKRYDVQGINAPFAGIDIKIRASMGKVDYHHD
EHDGERTSFDFNQANVNRLEASHTPIHTPMGKFGVGVLYLTSKNSGLVPPYEDG
NQDQNTIILHNKTKTSVWFPEEYKPNDKLTVDAAARIEKOTITMDYOKDAIYQSLN
LGLATONBEDIIRPKLLDSGTLPKPKOTARSVAVGHLOLTPKHKLSLNLSHOERLPN
AGELYAHGMHLATNSFEINRPLNKEKSNINILGLTFQGDWDRYLGGVHYHDPNVF
LQTLQYQKGLRGMHDKLKYARIEYQAAAKYGFEDVNIQYQINDVHVALFGDYIRG
KLUNLPDKKGRDADYGNRPLIKOPDSHTPLPPKRLGMKLTANVANWISGFLFYRHF
QKDLANPERPTPAHNLNLGNYQHKPSHQAGSVQVFFNANLLNDRKFAHETFPFD
MPQMRNFMGLGANFKF"
BASE COUNT      817 a      583 c      571 g      675 t
ORIGIN
Alignment Scores:
Pred. No.:      5,87e-87      Length:      2646
Score:      1585.00      Matches:      338
Percent Similarity:      58.92%      Conservative:      141
Best Local Similarity:      41.57%      Mismatches:      272
Query Match:      39.27%      Indels:      63
DB:      6      Gaps:      13
US-09-936-377-2 (1-758) x AX394914 (1-2646)
Qy      7  LysProIleValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db      110  AAACCGCTTGGCTGTGCCAATATGGCAACTTTTCAATGCCAATGCTGGCAGAGCGCAAT 169
Qy      26  ---GluThrGluGlnSerValGlyLeuGluThrValThrValGlyLysSerA-rGPro 44
Db      170  TTAAAGGATAAGCAACCGCTCATTTTAGATGGGCTTTCGATCACCTCTTTAGCTGACCAA 229
Qy      45  ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
Db      230  AATACAGAGTTTGGGCTTAATCATTCATAAAACAGTCAGTGGCATCACAGTTTCAAAGAG 289
Qy      65  ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
Db      290  CAACTACAAACAGCAGCAACCCCTTAGCGCATGCCCTTGGCAGGTGAGTTTGGCGTTCAT 349
Qy      85  AlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyA-rGArg 104
Db      350  TCTAACCATTTTGGGGCGGTGCTCAGCCCCCATCATCTCGTGGCAGGAGGTAACGC 409
Qy      105  IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
Db      410  CTGAAATCTCTACAAAACGGTTTCAGAGTTTGGACATGTTGGGTTGTCGCCAGACCAT 469
Qy      125  AlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArgGlyProValThr 144
Db      470  GCATAGCGGTGACACACACATCGCAAAACAGGTGGAGATTGTGCGAGGCTCTGGTGCC 529
Qy      145  LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
Db      530  TTGTTGTAGCTCTGGCAACTCAGCAGCGGCTGCAATGTCGTGATGACAAATATCCC 589
Qy      165  GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
Db      590  AGCAAAATTTGCCCAAGCAAA---TTACAAGGTGATGTGACGGTGGTCTTAGCAGTGCAC 646
Qy      185  LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
Db      647  CGTGAATAATTAATCACCAGCGAGTCCGAGCCCACTCGGAGAGCATGTGCGAGTGCCT 706
Qy      205  ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
Db      707  GTTGCGAGGGCTGTCCAAACAGCAGCAGACTATATAAACGCCACGCTTTGACCGCATGTC 766
```

```
QY 221 -----ArgAsnLeu 223
Db 767 TTTAACAATAAATCATGATGATAAAGTCACTCAGCCAGAAATTCATCTATAAAGACACCTTA 826
QY 224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
Db 827 AGCATCTGCCAGACAGCCATGCCAATCAAAACGAGGAACGCT-TGGCGTGTCATGGGT 885
QY 244 TrpArgLysArg-PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPr 263
Db 886 TGGCAATCAAGGCTTTTGGGGGCATCGGTGAGCTTACGCCGAGACAAATATGGCCTGCC 945
QY 263 oAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIl 283
Db 946 CAACCAATTCACATGAATATGAAGATGTAGCGTCATGGCAATTTCTCAGTCGCGCTTACA 1005
QY 283 eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAs 303
Db 1006 ATACAAGCCATATTTGGTTTGTATCTCTTTTGTAGGAATAATGATGACTTAGAGTTTGA 1065
QY 303 pAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHi 318
Db 1066 CAATCAGGGCTTGAATGCCATACATGATGATGACCAAGCAGCAGCAACCATGCTCA 1125
QY 318 sAlaHisAlaHis-----AsnGlyLysProTrpIleAspLeuArgAsnLys 333
Db 1126 TGACCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1185
QY 333 sArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgVa 353
Db 1186 GCGTTATGATGTCAGAGGCAATCAATGCGCGCTTTGTGGCATGTGATAAATCCGAGC 1245
QY 353 lHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAlaValGluAsnPh 373
Db 1246 CAGCATGGGCAGAAATGCGATTTATCATCATGATGATAGATAGTGGGGTGAGAAGACCAGCTT 1305
QY 373 ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle----- 390
Db 1306 TTTTGATATCAAGCCAAATGTGGCGTCGGAAGCGCTCATAACCCCAATTCATACGCC 1365
QY 391 ----GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
Db 1366 GATGGCAGCTTTAGCGGGGTGTTGGGGTAGGTATCTCACCTCAAAAACAGCGGACT 1425
QY 409 u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
Db 1426 TGTGCCACCTCGTTATGAAGATGGCAATAAACAAGACACCCAAACATCTTGCACAATAA 1485
QY 424 nLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn---TrpAspAsnPheth 443
Db 1486 TAAACCAAAACAGGCGAGTGTGTTTGGTTTGAAGATCAAAACCAATGACAAAGCTGAC 1545
QY 443 rLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLe 463
Db 1546 CGTTGACCGCGCGCTCGCATTTGAGAAACAAACCATCACCATGATATGATATAACAGCG 1605
QY 463 uIle-----AspArgGluAsnTyrTyrLy 471
Db 1606 CATTTATCAGACTTAACTTAGGCTTAGCAACCGCTCATGAACCAACAGACATACGCTTTAA 1665
QY 471 sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
Db 1666 ACGATTGCTGGACAGCGGTACTTTAAACCCCAAAACAAACCGCAGCTCTTATGCTGT 1725
QY 489 uSerGlyAsnTrpTyrPheTrpProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl 509
Db 1726 TGGGAGCGCATTTACAAATTAACGCCCAACATAAATATCGCTGAATCTGTGTCATCAAGA 1785
QY 509 uArgProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
Db 1786 ACGCTTGCATAATGCTCAGGAATTTGATGCTCAGCGCATGCACTTGGCCACCAACTCGTT 1845
```

```
QY 529 eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549
Db 1846 TGAATTTGGCAACCGCTTTTAAACAAGAAAAATCAACAACATTTGATTGGGGCTGAC 1905
QY 549 YTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
Db 1906 ATTTCAGGGTGAATAATGGGATTATCGTCTTGGGGGCTATCATATTGATTGTGATACTA 1965
QY 569 rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerIleGluAspAspSe 588
Db 1966 TGTGTTTTTACAAACATTTGCGCAGTATAAGCAGGTTTTCGCTGGCATCGCTCATGATA 2025
QY 588 rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGl 608
Db 2026 AGACTTAAACACCGCAGCTATGAACAAGCAGCGCGGAAATTTATGATTTGATGTCACAA 2085
QY 608 uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGl 628
Db 2086 CATCGGTTATCAGATTAAATGATGATCATGTGGCGTTATTGGTGATTTATTTCTGTGG 2145
QY 628 YArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPh 648
Db 2146 CRAAGCTCACCAATTTGCTGCACAAAAGGCGCAGACCGATGCGTATGGCAACGCTCTCT 2205
QY 648 eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668
Db 2206 CATCAAAACAGCCAGCAGTATACGCGCAAGACTGCCACCAAAACGCGCTTGGCATGAAAT 2265
QY 668 uLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGl 688
Db 2266 AACCGCAATGTAATGCAAAATGCTCAGGTTTTTGGAAATATCGCCATACCTTTAAACA 2325
QY 688 nAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAl 708
Db 2326 AGATAAATTTGGGAAATTTTGAACGCCCAACCCAGCTCATAACTTGGTGAATTTGGGGCT 2385
QY 708 aAsnTyrArgArgAsnThrArgTyr-----GlyGluTrpAsnTyrTyrValLysAlaAs 726
Db 2386 TAACTATCAGCAACAGCAAGCCATCAAGCAGGCTCGGTTTCAGGTATTTTAAATGCTAA 2445
QY 726 pAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMe 746
Db 2446 CAATCTATTAAACGATAAAAGTCTTGTCTCATGAGACATTTTCCCGACATGCCCAAAAT 2505
QY 746 tGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2506 GGGCGAAACTTTATGCTCGGGGCAATTTCAAATTT 2542
RESULT 10
LOCUS AX067459 89047 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 34 from Patent WO0078968.
ACCESSION AX067459
VERSION AX067459.1 GI:12545079
KEYWORDS
SOURCE Moraxella catarrhalis
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Moraxellaceae; Moraxella.
REFERENCE 1
AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 34 28-DEC-2000;
INCYTE Genomics, Inc. (US)
FEATURES
source
1. 89047
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"
66874
unsure
/note="a or g or c or t, unknown, or other"
BASE COUNT 26501 a 17338 c 19060 g 26147 t
ORIGIN
```

Alignment Scores:

Pred. No.: 5,09e-85 Length: 89047
 Score: 1585.00 Matches: 338
 Percent Similarity: 58.92% Conservative: 141
 Best Local Similarity: 41.57% Mismatches: 272
 Query Match: 39.27% Indels: 63
 DB: 6 Gaps: 13

US-09-936-377-2 (1-758) x AX067459 (1-89047)

QY 7 LysProLeuValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
 DB 424 AAACCGCTGCGTGGCATATTGGCAACTTTTCAATGCCAATGCTGGCAGGCGGAAT 483
 QY 26 ---GluThrGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
 DB 484 TTAAGGATAAGCAACCGCATTTTAGATGGCGCTTCGATCACCTCTTTAGCTGACCA 543
 QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
 DB 544 AATACAGAGTTGGCGTTAATCATCAAAACAGTCAGTGGCATCACAGTTTCAAAAGAG 603
 QY 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
 DB 604 CAACCTACAACAGAGCAACACCTAGGCGATGCCCTTGGCAGGTGAGCTTGGCGTTCA 663
 QY 85 AlaserGlnThrGlyGlyGlyAlaserAlaProValIleArgGlyGlnThrGlyArgArg 104
 DB 664 TCTAACCATTTTGGGGCGGTGGCTCAGCCGCCCATCATTCGTGGGCGAGGAGGTAAACGC 723
 QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 DB 724 CTGAAATCTACNAACGGTTCAGAGTTCAGACATGCTGGGTGTCGCCAGACCAT 783
 QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
 DB 784 GCCATAGCGGTGGACACACACCTGGCAAAACAGGTGGAGATTGTCGCGAGGCTCGTGCC 843
 QY 145 LeuLeuThrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
 DB 844 TTGTTGTACCCCTCTGGCAACTCAGCAGCGGTGGTCAATGCTGTTGATGACAAATACCC 903
 QY 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
 DB 904 AGCAATTTGCCAGCAAA---TTACAAGGTGATGTGACGGTGGCTCTTAGCAGTGCCAAC 960
 QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
 DB 961 CGTGAATAAATAATCACCGCAGTGGCGAAGCCCACTGGGAGAGCATGTGGCAGTGGCT 1020
 QY 205 ThrGluGlyLeuThrArgLysSerGlyAspThrAlaValProArgTyr----- 220
 DB 1021 GTTGCAGGGCTGTCCAAACAGCAGCAGACATATAAAACGCCACGCTTTTGCACCGCATGTC 1080
 QY 221 -----ArgAsnLeu 223
 DB 1081 TTTAAACAAAACATGAAGATGATAACACTCAGCCGAATTCATCTATAAAGCACCTTA 1140
 QY 224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
 DB 1141 AAGCATCTGCAGACAGCCATGCCAATCAAAACAGCAGCAACGCT-TGGCGTGTCTGGGT 1199
 QY 244 TrpArgLysArg-PheTyrArgArgThrTyrSerAspArgAspGlnThrGlyLeuPr 263
 DB 1200 TGGCAATCAAGGCTTTTGGGGGCATCGGTGAGCTTACGCCGAGACAAATATGCGCTGCC 1259
 QY 263 oAlaHisSerHisGluThrAspAspCysHisAlaAspIleIleIleIleIleIleIleIle 283
 DB 1260 CAACCATTCACATGAATATGAAGATGTACGTCGTCAGTGGCATTTCTCAGTCCGCTTACA 1319
 QY 283 eAsnLysArgTyrLeuGlnLeuThrProHisLeuLeuThrGluGluAspValAspTyrAs 303
 DB 1319

DB 1320 ATACAAGCCATATTGCTGTTGTATCCTTTTGTGATGGAATAATGATGACTTAGAGTTGA 1379
 QY 303 pAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHi 318
 DB 1380 CAATCGAGGGCTTCAATGCGCATACATCATGATGATGATGATGATGATGATGATGATGAT 1439
 QY 318 sAlaHisAlaHis-----AsnGlyLysProTrpIleLeuLeuArgAsnLys 333
 DB 1440 TGACACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1499
 QY 333 sArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgVa 353
 DB 1500 GCGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1559
 QY 353 lHisLeuAsnArgAsnAspTyrHisHisAspGlyLysAlaGlyAspAlaValGluAsnPh 373
 DB 1560 CAGCATGGCAAAAGTGATTATCATCATGATGATGATGATGATGATGATGATGATGATGAT 1619
 QY 373 ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle----- 390
 DB 1620 TTTTGATTAATCAAGCCCAATGTGTGGCGTCTGGAAAGCCTCATACATCCGCCATTATACGCC 1679
 QY 391 ---GlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
 DB 1680 GATCGGCAAGTTAGCGGGGTGTTGGGTAGGTATCTACCTCAAAAACACGCGGACT 1739
 QY 409 u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
 DB 1740 TGTGCCACCTCGTTATGAAGATGCAATAAACACAGACACCCAAAACATCTTGCACATAA 1799
 QY 424 nLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn---TrpAspAsnPhTh 443
 DB 1800 TAAACCAAAACACGCGAGTGTGTTTGGTTTGAAGAATAACAAACCAATGATGATGATGATGAT 1859
 QY 443 rLeuGluGlyGlyValArgValGluLysGlnLysAlaserIleArgTyrAspLysAlaLe 463
 DB 1860 CGTTGACCGCGCGCTCGCATTGAGAAACAAACCATCACCATGATGATGATGATGATGATGATGAT 1919
 QY 463 uile-----AspArgGluAsnTyrTyrLys 471
 DB 1920 CATTTATCAGAGCTTAAACTTAGGCTTTAGCAACCGCTCATGAACACGACATACGTTTAA 1979
 QY 471 sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
 DB 1980 ACATGCTGCGACAGCGGTACTTTTAAACCCCAAAACAAACCAACCGCAGCTCTTATGCTGT 2039
 QY 489 uSerGlyAsnTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGly 509
 DB 2040 TGGGACGCAATTACAAATTAACGCCCAACATTAATATCGCTGAATCTGTGCGATCAAGA 2099
 QY 509 uArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
 DB 2100 ACGCTGCCAAATGTCTCAGGAATTTGATGTCTACGGCATGCACCTTGCACCAACTCGTT 2159
 QY 529 eGluValGlyAsnLysHisLeuAsnLysGluLysSerAsnAsnIleGluLeuAlaLeuGly 549
 DB 2160 TGAATTTGGCAACCGCTTTTAAACCAAAAGAAATAACCAACATGATTTGGGGCTGAC 2219
 QY 549 YTyrGluGlyAspArgTyrPheGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
 DB 2220 ATTTCAGGTGATAAATGAGGATTCGTCTTGGGGCTATCATATTGATGATTTGATAACTA 2279
 QY 569 rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerIleGluAspAspSe 588
 DB 2280 TGTGTTTTTACAAACATTTGCGCAGTATAAGCAAGGTTTGGTGGCATGCGTCATGATAA 2339
 QY 588 rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGly 608
 DB 2340 AGACTTTAAACACCGCAGCTATGAACAGCAGCGCGGCAAAATTTATGATTTGATGTCAA 2399
 QY 608 uileTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGly 628
 DB 2400 CATCGGTATCAGATTAAATGATGATATCATGTCGGCGTTATTTGGTGGTATATATCGTGG 2459

Db 1816 CATCATAAATTGGCAATAGC-----TTTGTGGATAATCGAAGTGGCAGTGTGGA 1766

Qy	242	LeuGlyTrp-----ArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGln	259
Db	1765	CTTCTCTGGGTGGGTGATAAAGGTTATTAGGTGTTTCGTATAGCCAAAGTAAAGACAAA	1706
Qy	260	TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspLleIleTrpGln	279
Db	1705	TATGGGTACTCGGCATTCACATCTCTACGATGAATACTATATGCAATGTTTATTATATCC	1646
Qy	280	LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp	299
Db	1645	GATGCACATTGGCGTAAGCCCTACTTAAACATTAATCCATTTTAAATGGAGAAACTGAC	1586
Qy	300	ValAspTyrAspAsnProGlyLeuSerCys-----GlyPheHisAspAsp	314
Db	1585	ATTGATTATAATAATCCCGGTATTGATTGTATAAAAAAGAGTGGCATAGCCATGGCGAT	1526
Qy	315	AspAspAlaHisAlaHisAlaHisAsnGlyLys-----	325
Db	1525	CTGTGTAAATCACGGTCAATGCTCATCGGAAATGGGCAACATAGTCACGATCATCACGCG	1466
Qy	326	-----ProTrpIleAspLeuArgAsnLysArgTyrGluLeuLeuArgAlaGluTrpLys	342
Db	1465	CATGCCGATCGGCATATTCGTTTAATACCAACGTTGGGATTTGGTGCGGAATGGAAA	1406
Qy	343	GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHis	362
Db	1405	AATCCAGTAAAGGCTTAGATAAAGTCCGTTTTTCGATTGCCAAAGTTGGTTATCGCCAC	1346
Qy	363	AspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArg	382
Db	1345	GATGAAAAATCAGGCGGATCTCAGACAAATTCATTTAAAAATAAAGGCTATTTCAGCAGT	1286
Qy	383	IleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeu	402
Db	1285	GTTGAAATCTCTTATCAGCCCATTTGGGCGCTCAGTGGGTAAATGTTGTTTAACTCATGTA	1226
Qy	403	GlyGlnLysSerSerAlaLeu-----SerAlaThrSerGluAlaValLysGlnProMetLeu	421
Db	1225	TATCAAGATAGCTATGCGTTTAGATATAATCATACACTTTGATATCATGTAACAAATCTCCTT	1166
Qy	422	LeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAspAsn	441
Db	1165	TCGTGATCATACTACGGCTCAGCAAAAGCCTCTTTTAAATGAGCATGTGTAATTAGGTAAA	1106
Qy	442	PheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLys	461
Db	1105	TGCAATTGTATTTGTTGGCGGTGGAAAAACAACGCAATCCGATGAATAATCATTTT	1046
Qy	462	AlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeu---GlyAlaHis	480
Db	1045	AATGTGCCAAAAGATGAG-----CACCCACCGAAGAGCTGACTCGACCACAT	998
Qy	481	ArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLys	500
Db	997	AAATCGAAAGCCATTCTTATGCTTTATCGGCAAAATTATCAATTAATGAACAAACATCAG	938
Qy	501	LeuSerLeuThrAlaSerHisGlnArgLeuProSerThrGlnGluLeuTyrAlaHis	520
Db	937	TTTAAATGATGTTTTCGCATCAGAAGCTTTCCCAATCGCGAAGAGCTATTATGCGCAT	878
Qy	521	GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg	540
Db	877	GGAAAAACATTAGCCACCAATTCTTTTGAAGCGGGAATAAAAAATCTGCACAAAGAACGC	818
Qy	541	SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAla	560
Db	817	TCGAATAATGTGCAATTAGTTTGGGGGTATACAGGGGAAAAATTAGGATCAAAATTAAGT	758
Qy	561	LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsn-----	576
Db	757	GGCTACTATCAACAGTTTCTTCACTATAATTATGCGGCTATTTTTAATTAACAAAAACAAC	698
Qy	577	-----AspGlyArgGlyProLysSerIleGluAspAspSerGluMetLys	591

```

Db      697  TGTCCGTGGCGACCTAAATCCCGTTGTTACGTTTCATTAAGCGCATGATTATCCATTACGC 638
Qy      592  LeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyLuleTyrPhe 611
Db      637  TTATATCGCTACATCAGCGAAGCAAGATTTATGGTTGGAAGCGGAGATGAGCTAT 578
Qy      612  LysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyLeuLys 631
Db      577  CAATTTCTCTCAACACAGATGTGTCCATTTTGGTGATTTATGTCGGGGTAAGTTAAAA 518
Qy      632  AsnLeuProSerLeuPro-----GlyArgGluAspAlaTyr 643
Db      517  GACCTTCGCTCTCTCCCGATTGGATATAAATATATCTATAACGAAATTTATGACATGTT 458
Qy      644  GlyAsnArgPro-----PheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
Db      457  GGTGTACACCAACCAACGGTTGGGAAAAACAACCTGATGATGATGACCTCGTATGCACCA 398
Qy      662  AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
Db      397  ATGCGTTTATAGGCATAAAAATGGAATCGTATTTTGACAACGGCATAAAGTTTCAATACGCAG 338
Qy      682  TyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHis 701
Db      337  CTTTATCGAGCTTTTGCAAAACAAGAGTGGCCGACCTAGAACACCAACGAGGGGCAT 278
Qy      702  HisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyLysTyrAsnTrp 721
Db      277  ACCATGTTGAATCTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 218
Qy      722  TyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSer 741
Db      217  TTTGCTAATGTAATATGCTGAATAGTCTAATAGTCTAATAGTCTAATAGTCTAATAGTCT 158
Qy      742  AspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db      157  TATATCCGCAAGTGGTCTAGGCTTAATGTTGGAATGAATTTAAATTC 107

RESULT 12
AE006179/c LOCUS
DEFINITION Pasteurella multocida PW70 section 146 of 204 of the complete genome.
ACCESSION AE006179
VERSION AE006179.1
KEYWORDS
SOURCE
ORGANISM Pasteurella multocida
REFERENCE
AUTHORS May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and Kapur,V.
TITLE Complete genomic sequence of Pasteurella multocida, Pm70
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE 21145866
PUBMED 11248100
REFERENCE
AUTHORS Zhang,Q. and Kapur,V.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
FEATURES
source Location/Qualifiers
1..11375
/organism="Pasteurella multocida"
/mol_type="genomic DNA"
/strain="PM70"
/db_xref="taxon:747"
157..465
/gene="cyay"
/feature="PM1425"
/feature="PM1429"

```

CDS

157..465

/gene="cyay"

/codon_start=1

/transl_table=11

/product="Cyay"

/protein_id="AAK03509.1"

/db_xref="GI:12721802"

/translation="MNVTEPHQNTIEQVWAKIEEQLEAQDCVDCDCTQGSVFSITPQDR
SQVWINKPELLELWASRVGFFHRYQNNWNTSNDGKDFWACLEACAAGHEI VHS
"

gene

431..1351

/gene="PM1426"

431..1351

/gene="PM1426"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK03510.1"

/db_xref="GI:12721803"

/translation="MPLMVKLFTLVKEVIRMYKAGINLYKKLILAVLGVFVSHGTQ
ANESDARPYKSAEVLTKSDAFIGLLGYBENYLMGTYSNRHFLKREKTKQDEIKFKI
SLALPLARGILLGNNSVLAASVTKSMFQLSNVDSSPPRETYEPQLFLANKTQYSLP
FGWTLQDVEGINSQNGRDAEKLRSWNKLYVRAISAIRKONWVEIKPMWRKIPKAK
NDNPDIKTKRGHEDVALGYIYHDHDFKLSHYNPI SNKGLEASYSYPTITKIRFYT
QYNGYGESLIDYQORIQRIGIGISLNNVP"

gene

1341..3239

/gene="recQ"

/note="synonym: PM1427"

1341..3239

/gene="recQ"

/codon_start=1

/transl_table=11

/product="RecQ"

/protein_id="AAK03511.1"

/db_xref="GI:12721804"

/translation="MFSSLLSKDNKSAVDPEHPIPKOTALDVLHVAHVGYQSFRRGQE
EVIDATLMGKDSLVIMATNGKSLCYQIPALCFEGTLTIVISPLISLMKQDQVQLIANG
IEADYLSNSQDTPTEQQVQNKMSGLTKLYVSPKVMWTTTFHILSHKLNQSPHVAIDE
AHCISQWHDTPREYTOQLGGLKSCFHPALMALATADHATROQILRHNLNQSHPVYI
GSFDRPNIRYTLVEKFKPEBOLCHFLVGOKGSGI IYCNSSKVERIAESLRNKGVA
QAHAGLETQREQVQAFQORDNVQVATIAFGMINKNSNRFVVFHDLPRSESYY
QETGARDLLDPAEALVFYFPADYANLHLKLEKPEQFOIEALKQAIGFAESQT
CRRLVLNLYFGEHQKQPCNCDICLDPKQYDIDAKVMSTTIRIGQRFGVHYIA
VRLGNSQKIKNQHEQLSVYIGKDKSKHEWQSVIRQLIHGLIKOVDFHFNATLQ
TENAKPILRGEQPLSLAMPRISSLTSVAPQRYAIAOYDKLFLRLKQIAADKEN
IPAVIVENDATLQEMAOYQPTTKAEMLAINGVGATKFERFAQPFQMLIQHKKVLTQH
EPPLSLSS"

gene

complement(3338..5755)

/gene="PM1428"

complement(3338..5755)

/gene="PM1428"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK03512.1"

/db_xref="GI:12721805"

/translation="MISRGCKVNRKFFAVLMMCCIPQVWANTEKKQIVFLDEISVESK
GAAPRSDPLSGLPKONDILVSKOKLTGSSTLGNALAGELSVHNSQNGGSSAPVVRG
QEGVRLKILONGSDVLDMSQLSPDHAIQVDTLLABQVEIVRGASTLLYANASPAVIN
VWKRPTQLPKGYEVDFNTRYNINSHEKLVTAALTFTGLGKHLALRVEELRGSNNY
HVFAFKLDKTLNTPVDQNTKSGNYGVAFIGERYGVGFAYNLKREKYLGLPHGHKLD
SCAHLIWSGNVRNDYILGYLPHLMDTDLVNTFHCGSNHMDGKSHNHPYGHNDH
SIAGPLIDSYAKRYDIAEVKQPMKATEKLSYSETRYKHDEKGNIAVLPKNQNY
NLRAVEIHTPIAGLSVIGAOYTOTSSANI PRIAPCSNNASDPCHKKQORDPSKITK
GDRKSWALIENTOSQSPFAIEQLRWQDPEIGVTEKQRIDIEADRLMFKVRKRL
EGCDPSFFSPSGCRQGSYPADPASTHIDRATSYGAI SNWMTPTDLSLTSHNER
HPTPMELIYHGKHLATVSPFHGNNLKEYSNWEVGLAYIGDKLSYKNAVYNDPKN
RIFNQTLNKSGLNSLRNQSARKYQVEGRIDYALTPELHMLGFGYKGLYDLPP
TYRVDHANSLEPVTQDAPRPVPMRLGLFRVNMWMTESLTSLSLEYTYVQCKVAP
LENQTAAYSLNLNIGVDYSRQIAGVNYQLFQANNVLRKTYSHTSFLFPFQMGKRVNT
LGLNIHF"

gene

complement(5967..6779)

/gene="PM1429"

Db	3940	TA	CTTGGTGCACAACTGAAGTTAATAAGTGAATGATATATACAAATGATTTTAAAAATCGG	3988
Qy	570	ile	TyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu	589
Db	3880	AT	TTTTTAATCAAACTTGAAC-----AAATCCGGTAAT	3848
Qy	590	Met	LysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIle	609
Db	3847	TT	ATCTTTGAATCGTTAATAATCAATCAAAGCGAAATATATATGGTGGGAAGCGCGTATT	3788
Qy	610	Tyr	PhelysProThrArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArg	629
Db	3787	GAC	TATCGCTGGTACACCTGAACTGCACATGGGACTTTTGGTGATTATGTCGGTGGAAA	3728
Qy	630	Leu	LysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgPropheIle	649
Db	3727	CTG	TATGATTATTCGGCCAAACGTACCGTGTGCGATCATGTGGCGAATAGCTTAGAGCCGTT	3668
Qy	650	Ala	GlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLys	669
Db	3667	ACT	CAGCTGATCAAGACGCCACACGTGTCGCCAATGCGTTTAGCGTTTCGTGGAAT	3608
Qy	670	Ala	SerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsn	689
Db	3607	ATG	GAGATGACTCAGAGCTTTAAACAAGTTCACCTCGAATACACTTACGTTTATCAACAAAG	3548
Qy	690	Lys	LeuAlaArgTyrGluThrArgThrProGlyHisMetLeuAsnLeuGlyAlaAsn	709
Db	3547	AA	GTCGCGCTTAGAAAAATCAACACGCTGCATATAGTTTATTAATATCCGAGTGGAT	3488
Qy	710	Tyr	ArgArgAsnThrArgTyrGlyGluTyrPAsnTyrTyrValLysAlaAspAsnLeu	729
Db	3487	TATT	CAAGCCAAATAGCTGGGGTAAATATATCAATTATTTGTTCAAGCAATATATGCTTA	3428
Qy	730	Asn	GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSer	749
Db	3427	AA	CGTAAAGTTTATCTCATCTCTCTTTTACCCTTTGACCGCAGATGGGGCGTAAAT	3368
Qy	750	Phet	hrGlyValAsnValLysPhe	758
Db	3367	GTG	ACTTAGGATTAACATCCATTTC	3341
RESULT 13				
AE006112/c				
LOCUS				
DEFINITION	AE006112	10166 bp	DNA	linear BCT 08-MAR-2001
ACCESSION	AE006112	AE004439		
VERSION	AE006112.1	GI:12721038		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL	</			

```

gene      complement (75.. .1130)
/ gene="modC"
/ note="synonym: PM0742"
CDS      complement (75.. .1130)
/ gene="modC"
/ codon_start=1
/ trans1_table=11
/ product="ModC"
/ protein_id="AAK02826.1"
/ db_xref="GI:12721039"
/ translation="MLKINVKKQGLQALBANLOIPARGVTALFGLSGSKTSILNIV
SGLVHPDEGYSINERVLVQSKAVCIPAYORHIGYVFDARLPFPHYTVKGNLCYGIK
KIDIAKPDIDYILGIGIHLKRYPTITLSGGEKQKVAIGRALLTQPEILLMDPFLSALD
LPKRELSYLETTSKIDIPILYVTHSIBELLARBYVLLDEGKVPADRLESWE
NPLFLPKLEDEQAVLSLPIHLNNTGVYVLTALGLOQQQVWIKAQAEVGENIRICIK
GSDVLSLTQPEKTSIRNLLHGKYRIVERENVQIEIDKPIMATISKWALEDLA
LQLGQVPAQIKANSVL"
/ complement (1117.. .1845)
/ gene="modB_1"
/ note="synonym: PM0743"
CDS      complement (1117.. .1845)
/ gene="modB_1"
/ codon_start=1
/ trans1_table=11
/ product="ModB"
/ protein_id="AAK02827.1"
/ db_xref="GI:12721040"
/ translation="MFTEFLTSLGTTTETQAVKLSLSVSNAILNSLPVAILMAWLL
ARKNYGKLSVGLIHPLVLPVPIVGVYLLVAMGRNGFIKLYLSWFGFSFAPSWNG
AVLAARVAPLVPVRAIRLSIESIDMKLEQAOITLGASANEVFETTLPLSLPGLLAG
VVLGPARSLGFGATIAVSNIAGETQTIPLMTFTTQTPGAERQARLCVLAILLSL
ISLISEWLAKMKQKQCGSHAKD"
/ complement (2077.. .2841)
/ gene="modA"
/ note="synonym: PM0744"
CDS      complement (2077.. .2841)
/ gene="modA"
/ codon_start=1
/ trans1_table=11
/ product="ModA"
/ protein_id="AAK02828.1"
/ db_xref="GI:12721041"
/ translation="MLKLMITVSLATACAFSLSAQAKVTVFASASMTNALNIOAED
YKVKPDDELVPSPASTLAKQIEEGAPVDIFVSANTKMKHLSDKGLTKDTEKVL
VGNELVLTPASOLKEVDVAKGEMVAELKDSFLSCGDDPHVAGQVAKAELTKLW
DKVEAKLARGQDVRAALALVERAERAPLIGIVYSTGKVGKVAIVGVPPADYKAVEIP
VALLKDRDNAETREFNLDESAAKKVLVSYGFSVK"
/ complement (3009.. .5786)
/ gene="PM0745"
/ note="synonym: PM0745"
CDS      complement (3009.. .5786)
/ gene="PM0745"
/ codon_start=1
/ trans1_table=11
/ product="unknown"
/ protein_id="AAK02829.1"
/ db_xref="GI:12721042"
/ translation="MVNIIYFLYKKWIFLQNNIFMLNLSLTFVFKVKDQVETVKV
TNKLIKQITLPLPWASAVAFTHVIYAQTEMLDEIVVSGAOPHLAGSAIEHYQAIQNOVI
KBERLOQSATILCNALAGELGVHSPFGGASAPILRQEGVRIKILQNGLDVMSA
ISPDHARADSILQAQVELVGRNSTLYSSAAGVNVNDRIFTPAIVGEGEPIF
TRFDVASTGSTAGITFLRHPHLALRLGLKRYTHRYVFAFKSGGGETIRYLPDSHN
RQSGVTIGSWT KDQSYLGVSYSERDRYGLFGHNHRYDRCKSHVYDAAARPELGKGY
LTPYPLADDTDIPVAHLDCGICGIDNDPSSHDDPFGHEHDHSHGGPWRHLGRRFD
LRGSDSPAMLDKVKGSFYADITHYVHSGQAGTDKDRDSDFTIERERKKAENKRGK
AAGIYKNSGVNGLKLEFYHPTPIVGLSGVGVQVSEYKTAI LAPLGSIGIKHOHLVPNTQ
KQASFFAVENYVDDFIPEVGARVDKQRIPIKYDQHVNLNAHKKEGDHPDLSHKEKA
VSYLGSDWMLFHPNTRIGLTSRNERLPTPMELIYHGHLATNSFEHGNKNLRKESNN
NTELGPYDTKWDKLYSLYQNKFRNYINEDLARYGNAFLFRYTQARAFKHGEAEAL
NFRPTDYQVTLFGDYGRFLDPEYQGRFYQGYIAYDEEGLIQANWEKQPYRIEG
IKRNEADRPVPPARLMRLSNVTGTHLSFFADYTYVFSQQTASSLSIKPRALEAS
DFDIDDGTGNLLKGIDRDKYNRTQPTIADLSADKADPSALEAIRLRENEKLPAX
IEKIQDPSKGHLLNIGVNYQRIQIHLDSYGVFSVNNVLRQVTVHITSYLPYVPMQ
RNIVLNFVGK"

```

```

gene      complement (6825.. .7455)
/ gene="mode"
/ note="synonym: PM0746"
CDS      complement (6825.. .7455)
/ gene="mode"
/ codon_start=1
/ trans1_table=11
/ product="Mode"
/ protein_id="AAK02830.1"
/ db_xref="GI:12721043"
/ translation="MTLSEILLTIKQQQLFVDPKRIIRLLKEIKQCSINQAAKNAK
VSYKAWHLAMNRIKPKLLERNVGNKSGGTSLTVYVQRLLOLYLLSEHTQOKAF
HILQENPLDSVLHATFARFQSSARNOFGDI VALRHENIHCFVAQIKGLAFLIT
VATIKSQRLKLTGKVMKIKAPVWKVHRKEPEGVNTFLVNNKTEITDKGDVEVI
LTSINKHDEMEFCATLNKAENLPDEQIWIISIDPEQVILATLY"
/ complement (6868.. .7455)
/ gene="PM0747"
/ complement (6868.. .7455)
/ gene="PM0747"
/ codon_start=1
/ trans1_table=11
/ product="unknown"
/ protein_id="AAK02831.1"
/ db_xref="GI:12721044"
/ translation="MKISHFFACLDRLIQRWSLMRNLEKENLAHSHQVAFVAHAL
AVIKFYQCHLNPRIAMVNYHDTSEIFTGDLTPPKYFNPKITQAKETENRAEA
HLIALLPKELQDAFAPLDSTQPSAQERHVVKQADLI CAIVKAQFELENGNOEFKAAK
ARLETLMKTHSDMDRYFIEVPIPSFGKPIDETLL"
/ complement (7487.. .8515)
/ gene="degs"
/ note="synonym: PM0748"
/ complement (7487.. .8515)
/ gene="degs"
/ codon_start=1
/ trans1_table=11
/ product="Degs"
/ protein_id="AAK02832.1"
/ db_xref="GI:12721045"
/ translation="MIKKLPHSVMIGLVCAAVILVPSHVNQFSLKAPSPMSFKDAVR
VASPAAVNVNYSFSAEQLVGNLGSVIMSKDGYILTNKHVIQONADQIVVALQN
QGLDASLTGSDTLTDLAKIRADNLSTIPQNPQPMQHVMDVALAIGNPYNIGQVQ
QGISAIAGNAGDSVGRNQFTQTDASINRSGGALINSVGLIGINTLSIGKTSGE
IARGLNPAIPELANDVMKKIIRDGRVIRGYQVSDILFNNQGGHLEKGLITNVN
PEPGACKAGKAGDILKVGDDVDAISAPQMMQVINSNTRPNTRLNVVVISRLKILIEPV
IIEEYPLN"
/ complement (8512.. .9642)
/ gene="ribD"
/ note="synonym: PM0749"
/ complement (8512.. .9642)
/ gene="ribD"
/ codon_start=1
/ trans1_table=11
/ product="RibD"
/ protein_id="AAK02833.1"
/ db_xref="GI:12721046"
/ translation="NMTFTAQDVVYMQMALELAERKQFTTTPNPSVGCVLVKDGIIVG
RGFVRAGAEAAEVAERAGNARGATAYVTLEPCSHFQRTPPCAQGLIDAGVKKV
SMCDPNPOVAGRGLSMLDAGIESAVGLIAEAKWLNRCFLKRMRTGRYVQLKWAM
SLDGRYAMANGESKMTIGELARADYQAEKAKASAILSTSATVLADPELLNVRAQPLD
KVRADYTAQRLQRFVRLDSQHRVQPSHRLFSLSPVWLVSQTARSLLLEFPFACQEI
QLSASNTADLMALGELQQLINTLWIEAGATLAGALTEAKLVDELIIYIAPKLLGDAA
RGICKLPHLSTLAEASLWQLHSTKLGDDLLKTIYHPRIETPT"
/ complement (9642.. .10091)
/ gene="PM0750"
/ complement (9642.. .10091)
/ gene="PM0750"
/ codon_start=1
/ trans1_table=11
/ product="unknown"
/ protein_id="AAK02834.1"
/ db_xref="GI:12721047"
/ translation="MECPFCSTETKVIDSLYSEGVOYRRRECCNCHERTFTPTA
ELIIPKVNKDGTRFENBEKRRGIQHALEKRPVSENVEKAISYIIHRLRSTGRBE

```


645	-----AsnArgProPhe-----	648			
QY	649	-----IleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPhe			
DB	3572	GAAGGCATTAAACGTAATGAGCGTATGCGCTCGTGTGCCACCGCAGCTTTAGGTATG			
QY	667	HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPhe			
DB	3512	CGCTGAGTGGCAATTAACGGAACATTATCTTCTTTCGCAATACATTATGCTCTTT			
QY	687	AlaGlnAsnLysLeuAla-----			
DB	3452	TCTCAACAAATAACTGCTCTTCTTTTCAATTAACCGCGGAGAGCGTTAGAACGAGT			
QY	693	-----Arg			
DB	3392	GATTTTATTGATGATACAGCGCAAAATTTATTGTTAAAGGTATTGACCGGACAAA			
QY	694	Tyr-----			
DB	3332	TATAATCGTACGGTACGCTATCGCAGATTAAAGTGCTTCTGATAAAGCGGATTTGAA			
QY	694	-----			
DB	3272	AGTGCCTGGAAGCGATTCTGTTTAAAGAGAGAAATAATGAAATAATCCTGCTAAAAATAGAA			
QY	695	-----GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg			
DB	3212	AAATTCAGAGATCCGAGTAAGGCGCATCTTTGTTAAATATAGTGTAAATATCA			
QY	712	ArgAsnThrArgTyrGlyGluThrAsnTyrValLysAlaAspAsnLeuLeuAsnGln			
DB	3152	CGTCAAAATTTGGCATTTAGACTATTCTGTCGGTTTCAGTGTCTCAATATGTTATAACCAA			
QY	732	SerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThr			
DB	3092	CGTGTATTATGTTACACCTCTTATTACCTTATGTCGCACAAATGGGGCGCAATTATGTG			
QY	752	GlyGlyValAsnValLysPhe			
DB	3032	CTGAATTTTGGGGTAAATTC			
RESULT 14	AX394912	2942 bp	DNA	linear	PAT 18-MAY-2002
LOCUS	Sequence 49	from Patent WO0218595.			
DEFINITION	AX394912				
ACCESSION	AX394912.1	GI:21065985			
VERSION					
KEYWORDS					
SOURCE	Moraxella catarrhalis				
ORGANISM	Moraxella catarrhalis				
REFERENCE	1				
AUTHORS	Loosmore, S., Wang, J., Bradley, B., Ochs, M. and Yang, Y. P.				
TITLE	Moraxella polypeptides and corresponding dna fragments and uses thereof				
JOURNAL	Patent: WO 0218595-A 49 07-MAR-2002;				
FEATURES	Aventis Pasteur Limited (CA)				
source	Location/Qualifiers				
	1..2942				
	/organism="Moraxella catarrhalis"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:480"				
	101..2842				
	/note="unnamed protein product"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAD32082.1"				
	/db_xref="GI:21065986"				
	/translation="MIKKPLVCAISATFAMPVADNTKLGEPTTTTLKGVLYSSQTNQ				
CDS					
Alignment Scores:					
Pred. No.:	6,44e-64	Length:	2942		
Score:	1205.50	Matches:	290		
Percent Similarity:	46.25%	Conservative:	142		
Best Local Similarity:	31.05%	Mismatches:	303		
Query Match:	29.87%	Indels:	199		
DB:	6	Gaps:	18		
US-09-936-377-2 (1-758) x AX394912 (1-2942)					
QY	9	IleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGlu	28		
DB	83	ATCATACCGAGGATATCATGATAAAAAAACCACTTCTGTTGCGGATATCGGCCACCTTT	142		
QY	29	GlnSerValGlyLeuGluThrValThrValValValValValValValValValValVal	48		
DB	143	CGCATGCCAGCGGATGAGAGATATACCAAGCTGGTGAGAGAGCAACACCACTTTAAAG	202		
QY	49	GlyLeuLeuHisThrSer-----	54		
DB	203	CGTGTATTGTTAAAGTTCGCAACCAACCAACCAACCAACCAACCAACCAACCAACCA	262		
QY	55	ThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGly	74		
DB	263	CAATCCAGTATCTTACGCTTTCAAAAGATATAATATAATATATATATATATATATAT	322		
QY	75	AspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAla	94		
DB	323	AATGGTTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	382		
QY	95	ProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThr	114		
DB	383	CCTGTTGTGCGAGGCGAGAGGCTGCTCTTAAGATTTTACAAATATGGAACCTGATGTG	442		
QY	115	GlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGln	134		
DB	443	ATTGATGTTTCAAT	502		
QY	135	GlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGly	154		
DB	503	AAAGTTGAGCTTGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	562		
QY	155	LeuValAspValAlaAspGlyLysIleProGlyLysMetProGluAsnGly-----	171		
DB	563	GTGATTAAATGTTGTTGATGACCGTATCCGAATCGTATGCTGCTGCTGCTGCTGCTG	622		
QY	172	---ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSer	190		
DB	623	AAATTCGAGGCGAGAGCGATGCTTCGATATACACAAACCAACCAACCAACCAACCA	682		
QY	191	GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArg	210		
DB	683	GCAGGGGTGAGCTTTGGGGTAGGAGATCGCATTCGGGTTCGGGTGGAGGGCTTAAAG	742		
QY	211	LysSerGlyAspTyrAlaValProArgTyr-----ArgAsnLeuLysArgLeuPro	227		

```

Db      743  GAGCGTGATGACTATCAAGTTCCCAATTTTCAGGCGAGATCGCATGTAGATTATGGCCA 802
Qy      228  AspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGly----- 243
Db      803  GGTAGT-----GCAATAACTCTACCGTTGGCATGATTGGCGTGTCTATATT 850
Qy      244  TrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPro 263
Db      851  CATGATAATGGGCATATCGGTGCTTCTTATAGCCACCGTAAGATCGTTGATGATCCCA 910
Qy      264  AlaHisSerHisGlnTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIle 283
Db      911  GGCATATCCACTCGCGACGCCACGAGAGCAATTTATCAATGGCATAATATCAAAA 970
Qy      284  AsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAsp 303
Db      971  TCCAAATTATTATTTACCATTATTCCTCATTTGATGAGGATTCAGATATTGAT---GAC 1027
Qy      304  AsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAsn 323
Db      1028  AATCCTCATACGCAATTCGCCCAACACGAGACCATATCGTGAGCATATCCACG 1087
Qy      324  Gly-----LysProTrpIleAspLeuArgAsnLysArgTyr 335
Db      1088  GCGCGTCCCATCAATCATGAACATCATTCGCTTGGATTGATATGAAACCAATCGCTAC 1147
Qy      336  GluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu 355
Db      1148  GACATTCTGGCGAGGTATATCGCCCTATTCAAGTTTGGATAAATTAAGCTTAAGCTT 1207
Qy      356  AsnArgAsnAspTyrHisAspGluLys---AlaGlyAspAla----- 369
Db      1208  ACTTATGCAGATTATTATCATGATGAAAGATGCTGCCAATGAGCAAGACCCAAACAAT 1267
Qy      370  -----ValGluAsnPhePheAsn 375
Db      1368  CACAAACCTTCTGAGCGTGATACAAACGCTGGATAGGTCATGCCAGCTCTATTTTACA 1327
Qy      376  AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly 395
Db      1328  AAAAAAGCGTTAATGTGCTGCTGGAGTTATATCATACCGACCAACGCTTATCTGGG 1387
Qy      396  SerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAla----- 411
Db      1388  GTATTGGGTATGGGTATCAAAACCAAAATCTCGACGAGAGGCGGTATTTGCCAAGC 1447
Qy      412  -----ThrSerGluAlaValLysGlnPro----- 419
Db      1448  TATTTTCAATCAGAAGCAGATGCGCAAAAGCCCAAAAGTCAAAACATTAACCAATATCGT 1507
Qy      420  -----MetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAla 437
Db      1508  CTTTACTATTATAGTTCCAAATACCAATAAAGACCTTGGTATTTTGGACTTGAGCACTA 1567
Qy      438  AsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIle 457
Db      1568  AAGCTAAATCAATGACTTTTAAAGTGGCGATGCTCATGAAGACAAACCAACCAAT 1627
Qy      458  ArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyr----- 470
Db      1628  GAATATGACGACATTTACTTTGACCATGCTTTTGACGATTTATTTTAAAGTAAAGCACAGCTA 1687
Qy      471  LysGlnPro---LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
Db      1688  AAAGCACCTGATCATCTCTGATTTGACGACATATAACACATGCGCCTCTTATGCTGGT 1747
Qy      490  SerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu 509
Db      1748  AGTGCCTTATGGCATATTACGCCAAATCATCGATTGTCATTGACCTACTCACATAACGAA 1807
Qy      510  ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPhe 529

```

RESULT 15

AX067466

LOCUS

```

Db      1808  CGCATTCATCGCGCATGAGCTGATTATATCAAGCGGACATTTGGCGACGACTCTTTT 1867
Qy      530  GluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly 549
Db      1868  GAGCATGGCAATAAAACTTGTCTAAAGAAAATTCGGATTAATATGAGCTGGGTTTATG 1927
Qy      550  TyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
Db      1928  CATACAGAGATAAAGTCAGTCATATAAAGCAAGCACTTACTATACAAATTTTGATACTAT 1987
Qy      570  IleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu 589
Db      1988  ATTTTAAATGAGACC-----ATTGCCAAAGAGGCAAT 2020
Qy      590  MetLysLeuValArgTyrAsnGlnSerClyAlaAspPheTyrGlyAlaGluGluIle 609
Db      2021  TTATACATCAGACGCTATAATACAGACGCGCTAAGTTTATGTTGGAGGGTTCATTA 2080
Qy      610  TyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArg 629
Db      2081  ACTTACCAGCCAAATGCCAATCACAGTGTGATGTTTTTGGTGATATGGTCAAGGTAAA 2140
Qy      630  LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
Db      2141  ATTGGTGCATTATCTGATTAATTAAGGTAAAGCTT---GTATATGCTGGCAGAAAGTGGGT 2197
Qy      650  AlaGlnAspAspGln----- 654
Db      2198  TATTTTGGATGATATCATCAAGGATATGACCGTTGATGATATGTTGATTATGATCCGAT 2257
Qy      654  ----- 654
Db      2258  GGTGTTTGACTTGTGCTTAAACCGCTGAACAATGGGACAAAATTAACGATAATAAT 2317
Qy      654  ----- 654
Db      2318  GATTGTAGTACACCATTAATGTCTATAAAAACCGGTACACAACTCAGGCGAGGAGAT 2377
Qy      655  -----AsnAlaProArgValProAlaAlaArgLeuGly 665
Db      2378  TACGACCGTTTGGCAGCAAAATCAACTTACGACCCAGAGTACCGCCCGCGTGGGC 2437
Qy      666  PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
Db      2438  ATTCGTTGGCAAGGATTTTGGTGATCATTTGCTCTGCCAATGACAGAAATTTAAACCATGTG 2497
Qy      686  PheAlaGlnAsnLysLeuAla----- 692
Db      2498  TTTGCACAAAATAAAGTTTGCCACCTCAACGCTTGCCATTAAACCTCAATTCACGACGCA 2557
Qy      692  ----- 692
Db      2558  GAAGGTTGCCAACGCCATGAGAGTCATTTGCCGAATACGCGACTATGGCAGTGTATAACAAC 2617
Qy      693  -----ArgTyr-----GluThrArgThrProGlyHisMetLeu 704
Db      2618  CCTTTGATGATGATGATGATATATACAGAAACAAACGCGAGGATATAATTTGCTC 2677
Qy      705  AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrVallys 724
Db      2678  AATGTTGGCTAGATTATAACAATGATATGATGTTGATTATACGCTGCTCAATTGCT 2737
Qy      725  AlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPro 744
Db      2738  GCGAATAATTTACTTAATGAACAAATCATATTCAAACTCATTTTGGCGGTTGTACCG 2797
Qy      745  GlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db      2798  CAGATGGGCGTAATCTGACTTTAGTTTACGGGCTAAATTT 2839

```

269223 bp

DNA

linear

PAT 24-JAN-2001

```

DEFINITION Sequence 41 from Patent WO0078968.
ACCESSION AX067466
VERSION AX067466.1 GI:12545091
KEYWORDS Moraxella catarrhalis
SOURCE Moraxella catarrhalis
ORGANISM Moraxella catarrhalis
REFERENCE Lagace,R.E., Patterson,C. and Berg,K.L.
AUTHORS Nucleotide sequences of moraxella catarrhalis genome
TITLE Patent: WO 0078968-A 41 28-DEC-2000;
JOURNAL Incyte Genomics, Inc. (US)
FEATURES
    source
        1..26923
            /organism="Moraxella catarrhalis"
            /mol_type="genomic DNA"
            /db_xref="taxon:480"
BASE COUNT 77067 a 56596 c 57380 g 78180 t
ORIGIN
Alignment Scores:
Pred. No.: 1..99e-61 Length: 269223
Score: 1205.50 Matches: 290
Percent Similarity: 46.25% Conservative: 142
Best Local Similarity: 31.05% Mismatches: 303
Query Match: 29.87% Indels: 199
DB: 6 Gaps: 18
US-09-936-377-2 (1-758) x AX067466 (1-269223)
QY 9 IleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGlu 28
Db 209651 ATCAATACCGAGGGTATCATGATAAATAAACAACACATTTGTTGGATATCGGCCACCTTT 209710
QY 29 GlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSer 48
Db 209711 GCGATGCCAGCGTAGCAGATAATACCAAGTGGGTGAAGAGAGAGAGAGAGAGAGAGAG 209770
QY 49 GlyLeuLeuHisThrSer----- 54
Db 209771 GGTGATTTGGTAAGCTCGCAACAGAACCAATAACAGGTTTGTATCTAATGATTCACAAA 209830
QY 55 ThrAlaSerAspIleIleSerGlyAspThrLeuArgGlnLeuAlaValAsnLeuGly 74
Db 209831 CAATCCAGTATCTACGCTTCAAGAGATAATTAATATATCGTTCGCAACCTTGGGC 209890
QY 75 AspAlaLeuAspGlyValProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAla 94
Db 209891 AATGCGTTAAGTGGTGGAGCTTGGTATTCATGATTAACCCCTTTTGGTGGGCTTCATCTGCA 209950
QY 95 ProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThr 114
Db 209951 CTTGTTGTCGAGGCAAGAGGGTGTGCTCTTAGATTTTACAAATGGAAGTGAATG 210010
QY 115 GlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGln 134
Db 210011 ATTGATGTGTCATCAATATPCGCTGATCATGTGTGTGGCAGCATACATCTTTAGCGTCT 210070
QY 135 GlnValGluIleLeuArgGlyProValThrLeuLeuThrLeuLeuLeuLeuLeuLeuLeu 154
Db 210071 AAAGTTGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 210130
QY 155 LeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly----- 171
Db 210131 GTGATTAATGTTGTTGATCAGCGTATCCCGAATCGTATGCTGCTGCTGCTGCTGCTGCTGCTG 210190
QY 172 ---ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSer 190
Db 210191 AAATCGAAGGCGAGAGAGTGTCTTCATATACCAACCAACCAACCAACCAACCAACCAACCA 210250
QY 191 GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuThrArg 210
Db 210251 GCAGGGGTGAGCTTTGGGGTAGGAGATCGCATTCGGTTCGGTTCGGTTCGGTTCGGTTCGGTTC 210310
QY 211 LysSerGlyAspThrAlaValProArgTyr-----ArgAsnLeuLysArgLeuPro 227
Db 210311 GAGGCTGATGACTATCAAGTTCCTCCCATTTTCAGGAGATCGCATGTTAGATTATGTGCCA 210370
QY 228 AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly----- 243
Db 210371 GGTAGT-----GCAATAACTCTACCGTTGGCATGATTGGCGGTGCTTATATT 210418
QY 244 TrpArgLysArgPheTyrArgThrThrSerAspArgArgAspGlnThrGlyLeuPro 263
Db 210419 CATGATAATGGGCATATCGGTGCTCTTATAGCCACCGTAAAGATCGTTATGATATCCCA 210478
QY 264 AlaHisSerHisGluThrAspCysHisAlaAspIleIleTrpGlnLysSerLeuLeu 283
Db 210479 GGGCATATCCACTGCCAGCCCAACGAGAGCATTTTATCAATGGCATATATACAAAA 210538
QY 284 AsnLysArgTyrLeuGlnLeuThrProHisLeuLeuThrGluGluAspValAspTyrAsp 303
Db 210539 TCCAATTATTATTATACCATTTATCTCTCATTTGATGGAGATTTCAGATTATGAT---GAC 210595
QY 304 AsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAsn 323
Db 210596 AATCCTCATACGATTCGCCGCCACACAGAGACCATATCGTGAGCATATATCCACG 210655
QY 324 Gly-----LysProTrpIleAspLeuArgAsnLysArgTyr 335
Db 210656 GGGGTGCCCATCAATCAATCAATCAATCGCTTCGATGATGATGATGATGATGATGATGATGAT 210715
QY 336 GluLeuArgAlaGluThrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu 355
Db 210716 GACATTCGTGGGAGGTATATCGGCTTATTCAGGTTTGGATAAAATTAAGCTAAGCTTA 210775
QY 356 AsnArgAsnAspTyrHisHisAspGluLys---AlaGlyAspAla----- 369
Db 210776 ACTTATGCAGATTATTATCATGATGATAAAGATGCTGGCAATGAGCAAGACCCCAACAAT 210835
QY 370 -----ValGluLeuPhePheAsn 375
Db 210836 CACAAACCTTCTGCGGTGATACACAGGTGATGATGATGATGATGATGATGATGATGATGAT 210895
QY 376 AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly 395
Db 210896 AAAAAAGCGTTAATGTCGCTTGGAGTTATATATACACCGACCAACCGCTTATCTGGG 210955
QY 396 SerTrpGlyValGlnThrLeuGlyGlnLysSerSerAlaLeuSerAla----- 411
Db 210956 GTATTGGTATTTGGGTATCAACCCCAAAATATGCGAGCAGAGAGCGGTATTTGCCAAGC 211015
QY 412 -----ThrSerGluAlaValLysGlnPro----- 419
Db 211016 TATTTTCATCAGACGAGATGCGCAAAAGCCCAAGCAATTAACCAATATCGT 211075
QY 420 -----MetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAla 437
Db 211076 CTTTACTTATAGTTCCTCAATACCAATAAAGCGCTTGGTATTTTGGAGCTTGAGCAACTA 211135
QY 438 AsnTrpAspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIle 457
Db 211136 AAGCTAAATCAATGCTTTTAAAGGTGGCGATGCGTCATGAAAGACAAAAACCACAAT 211195
QY 458 ArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyr----- 470
Db 211196 GAATATGACGAGCATTTACTTGTACCATGCTTGTGAGTATTTTAAAGTAAGCAGCACTA 211255
QY 471 LysGlnPro---LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
Db 211256 AAAGCACTGATCATCTGATTTGACGACATATAAACAACATCCACCTCTTATGCTGT 211315
QY 490 SerGlyAsnTrpThrProGlnHisLysSerLeuThrAlaSerHisGlnGlu 509
Db 211316 AGTGCCTTAGGATATTACGCCCAATATCATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 211375

```

```

RESULT 16
AX052569          2742 bp      DNA      linear      PAT 12-JAN-2001
LOCUS              Sequence 11 from Patent WO0071724.
DEFINITION
AX052569          AX052569
ACCESSION
AX052569          AX052569
VERSION
AX052569          AX052569
KEYWORDS
AX052569          GI:12226785
SOURCE
AX052569          Moraxella catarrhalis
ORGANISM
AX052569          Moraxella catarrhalis
REFERENCE
AUTHORS
TITLE
JOURNAL
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BB)
FEATURES
source
1. 2742
/organism="Moraxella catarrhalis"
/mol_type="genomic DNA"
/db_xref="taxon:480"
BASE COUNT      844 a _536 c 600 g 762 t
ORIGIN
Alignment Scores:
Pred. No.:      9.59e-64      Length:      2742
Score:          1202.00      Matches:    291
Percent Similarity: 46.66%      Conservative: 142
Best Local Similarity: 31.36%      Mismatches: 301
Query Match:    29.78%      Indels:     194
DB:              6      Gaps:        20

US-09-936-377-2 (1-758) x AX052569 (1-2742)
QY 7 LysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGlu 26
Db 10 AACCACCTGTTGTGGC-----ATATCGCCACCTTTCGATGCCAGCGTAGCA 60
QY 27 ThrGluGlnSerValGlyLeuGluThrValThr-----ValValGlyLysSer 42
Db 61 GATAATACCAAGCTGGGTGAAGGCCAACCCACCCTTAAAGGGTGTGTTGGTAAGCTCG 120
QY 43 ArgProArgAlaThrSerGlyLeuLeuHisThrSerThr-----AlaSerAspLysIle 60
Db 121 CAACGACGACCAATACAGGTTTGTATCTAATGATTCAAAACAATCCAGTATCTACG 180
QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db 181 CTTTCAAAAGATAAATTAATAATATCGTTCGGCAACCTTGGGCAATGCGTTAAGTGTGAG 240
QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
Db 241 CTTGGTATTCATGTAACCTTTTGGTGGGTTCACTCGACCTGTTCGCGAGGCA 300
QY 101 ThrGlyArgGlyLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
Db 301 GAGGCTGTGCGTCTTAAGATTTTACAAATGGAATCATGTGATGATGATGATCAATA 360
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluLeuArg 140
Db 361 TCGGCTGATCATGTGTGGGACCGATACACTTTTAGCGTCTAAAGTTAGCTTGTTCGT 420
QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
Db 421 GGTCTGATACGCTGTTATATATGCTTGGCTGCGCATCGGCTGCTGATGATGATGATGAT 480
QY 161 GlyLysIleProGluLysMetProGluAsnGly-----ValSerGlyGluLeu 176
Db 481 GACCGTATCCGGAATCGTATGCTTCCCTAGTGGTGTCTATCCATGACAAATGAGGAGACG 540
QY 177 GlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGly 196
Db 541 ATGTTTGTATATAACAAACCAACCATGAAAGCTTGCACCTGCGAGGGGTGAGCTTGGG 600

```


Db	2716	ACTTTAGCTTTCACGGCTAAATTT	2739
RESULT 17			
AE012378		10947 bp	DNA linear BCT 23-MAY-2002
LOCUS		Xanthomonas campestris pv. campestris str.	ATCC 33913, section 286
DEFINITION		of 460 of the complete genome.	
ACCESSION		AE012378	AF008922
VERSION		AE012378.1	GI:21113832
KEYWORDS		Xanthomonas campestris pv. campestris str.	ATCC 33913
SOURCE		Xanthomonas campestris pv. campestris str.	ATCC 33913
ORGANISM		Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;	
		Xanthomonadaceae; Xanthomonas.	
REFERENCE		1 (bases 1 to 10947)	
AUTHORS		da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,	
		Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida	
		Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,	
		Carmago,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,	
		Chambergo,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L.,	
		Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S.,	
		Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,	
		Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite	
		Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,	
		Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,	
		Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,	
		Okurá,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,	
		Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,	
		Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos	
		Santos,M., Truffi,D., Teai,S.M., White,F.F., Setubal,J.C. and	
		Kitajima,J.P.	
TITLE		Comparison of the genomes of two Xanthomonas pathogens with	
		differing host specificities	
JOURNAL		Nature	417 (6887), 459-463 (2002)
MEDLINE		22022145	
PUBMED		12024217	
REFERENCE		2 (bases 1 to 10947)	
AUTHORS		da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,	
		Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida	
		Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,	
		Carmago,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,	
		Chambergo,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L.,	
		Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S.,	
		Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,	
		Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite	
		Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,	
		Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,	
		Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,	
		Okurá,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,	
		Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,	
		Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos	
		Santos,M., Truffi,D., Teai,S.M., White,F.F., Setubal,J.C. and	
		Kitajima,J.P.	
Direct Submission			
Submitted (28-NOV-2001)		Departamento de Bioquímica, Universidade de	
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,		Brazil	
FEATURES		Location/Qualifiers	
source		1..10947	
		/organism="Xanthomonas campestris pv. campestris str. ATCC	
		33913"	
		/mol_type="genomic DNA"	
		/strain="ATCC 33913"	
		/db_xref="ATCC:33913"	
		/db_xref="taxon:190485"	
		/notes="pathovar: campestris"	
		complement(85..501)	
		/gene="XCC2657"	
		complement(85..501)	
		/gene="XCC2657"	
gene		/note="identified by sequence similarity; putative; ORF	
		located using Blastx/Glimmer/Genemark"	
CDS		located using Blastx/Glimmer/Genemark"	
		codon_start=1	
		/product="transcriptional regulator"	
		/protein_id="AA041931.1"	
		/db_xref="GI:21113835"	
		/translation="MLSLRCTCFHLRRARRVQLYDHALAPVGLSLNEYSILRLRE	
		VQSIGLSTASLTGMDRTTLNLKPLIAGWITTORGEDAROKVRLSRSGRLLEA	
		RAQHWAQVOLETUFGAAPTAQLHATLARLDVALEHA"	
		3509..4738	
		/gene="XCC2660"	
		3509..4738	
		/gene="XCC2660"	
		/note="identified by sequence similarity; putative; ORF	
		located using Blastx/Glimmer/Genemark"	
		codon_start=1	
		/product="transport protein"	
		/protein_id="AA041932.1"	
		/db_xref="GI:21113836"	
		/translation="MSTSAAPRTRAHWGLLVTSAMLLTMGAROTTGLFVEPIHRQT	
		GIGTASTPALAIQOLYVGAVQPVFGEIAIDQRGPLVMIEGVLNAGLGQIFAPFLP	
		WGLIVSLVLAAGAAGSFSVLIGATAIRIAPERSFAAGLNAGSLGQIFAPFLP	
		QLTIGAQAQAKMTGLIALVSLITFLANPLIRVPVSATGAHTGAVAGEISLRACLAL	
		KDRSYWLHLGLFTFCGVHIATFLVTHLPCEVALGELPASVASVAGIALGLEFNAGSLTI	
		GRGERVRMKMLLTANTYASRAVMIGLIYIIAPPITPLTYLPAAGLGTFLWTATVPPTGAL	
		IKGLFPRYLGTFLGMLLSHQLGFFGAWLGGGLAHSNGYLMWYGDIVLVAAL	
		VNLPIREAVPVVRHAVA"	
		complement(4847..5866)	
		/gene="cxtB"	
		/note="synonym: XCC2661"	
		complement(4847..5866)	
		/gene="cxtB"	
		/note="identified by sequence similarity; putative; ORF	
		located using Blastx/Glimmer/Genemark"	
		codon_start=1	
		/product="conserved hypothetical protein"	
		/protein_id="AA041929.1"	
		/db_xref="GI:21113833"	
		/translation="MPLPTLRDLORFGATGSLCAVHCAPVLPLALAPSLGLSFWSL	
		GDGVELAIIVFVTLLGLFSLVLYGYRRHKALHALALLPGILLWLLGLLYDPLRHSSWP	
		HAVVMTLGGALVGIHVLNLRNLNHGHGHVDASCAR"	
		663..2729</	

```

/codon_start=1
/transl_table=1
/product="cytochrome D oxidase subunit B"
/protein_id="AA041933.1"
/db_xref="GI:21113837"
/translation="MFVHLSVELMQHLASAALAFVTYVVVLDTGLGVCMLEAANRS
DADRHLMAYSILPWNQNETWLVLLGGSSLLAMPESAIAIENATYPLVFWMLLALICR
GVALRYRLQGTDRRRRVLDVWFVASLLATFCQGVAGTLAGVAHRAGQYAGNRCR
WLQPPFLPGFAGVALLVGYLIGACMLWMRSBGQLQQRARHQARVLGIVLGALLTGVM
LARUTPYAQRLGDPAHSVLPGLTGMVAMYLIMSRSHAFLEFLFAVLGVVATVATG
VLAIYPMVLDPDITLAQAASPPASQSGVMLLAFAVLVPCTLASYSYGFYVRGKVKPR
GHAE"
gene complement(5882..7255)
/gene="gxtA"
/note="synonym: XCC2662"
CDS complement(5882..7255)
/gene="gxtA"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="cytochrome D oxidase subunit A"
/protein_id="AA041934.1"
/db_xref="GI:21113838"
/translation="MQASETHAALVYVGTIPMEPVGLHWGRSLGSAGAVIGPLMPFE
RTRRVMDVNYWKLPSLTAAVGVGTIPMEPVGLHWGRSLGSAGAVIGPLMPFE
VLVAFLEAGFLMGFKMKVGRHVHFATCVMAVGLSVSAEWILLSNSLHTPAGP
ARAQDNFLPDVDMRIIFNPSPFRMHLMSLAALLATAIAAGWAHHVDURDTNRGA
RIMSMGLWVIATVAPQLDLHGRLTEHQPKLAAIEGSRQPAPAGEBPELRF
AIPDAQRNHEILAIRLGSLYLRHNWTGTIAGREFPRSDFPVPPIVFPAFRAMVG
LGLLIAVGLMRSLRGYEARLVQTVVWMAPSMLMLCGWVTVTEVGRQPTTY
VGLLRTAHSTSPFLSPMWVSVAAILVLYPLVFGLLGLLRVLRLPHARENGSPSG
LADQTGN"
gene complement(7268..7753)
/gene="XCC2663"
CDS complement(7288..7753)
/gene="XCC2663"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA041935.1"
/db_xref="GI:21113839"
/translation="MAEPLDYDVKNYPPIGPFTAREELDRLETTLHQRLGILANDLVGA
ODKLAKVLIDGNOGPLNVQSYILAMAVSRIPPQQFYKVFAKGDAOLDQIARERA
QTAAQGEAPGLRGAYRLLEDQEAVNNAITPVLDGKFAAGLDKQVKPIPSAGSKPSE
S"
gene complement(7753..10725)
/gene="XCC2664"
CDS complement(7753..10725)
/gene="XCC2664"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="formate dehydrogenase related protein"
/protein_id="AA041936.1"
/db_xref="GI:21113840"
/translation="MQGHSCHTTPYDGKTLHGQADIPLIEFSGHGIDLHCITYPAL"

```

Alignment Scores:

Pred. No.:		2,63e-35	Length:	10947
Score:		745.00	Matches:	219
Percent Similarity:		43.02%	Conservative:	123
Best Local Similarity:		27.55%	Mismatches:	301
Query Match:		18.46%	Indels:	152
DB:		1	Gaps:	21
US-09-936-377-2 (1-758) x AE012378 (1-10947)				
Qy	1	MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuAAsnThrProLeu	20	
db	687	CGTGGCGGTCCACCTACGTCGGCCCTGGCGCAGTGTCTGATC-----CCCGCT	737	

Query Match:	17.80%	Indels:	149
DB:	1	Gaps:	21
US-09-936-377-2 (1-758) x AB011924 (1-11947)			
Qy	11	LeuSerIleLeuLeuIleAAsnThrProLeuLeuAlaGlnAla-	24
Db	1333	CTGGCCCTGGCAGCCTGCTGACCCCTGCTCCCTGCGCATCGCGAGGACGCTCCGCTCGCT	1392
Qy	25	-----HisGluThrGluGlnSerValGly 32	
Db	1393	GCCGATGGCGACAGCCCGCCAGTTCGGACACACGCCAGGAGCGAGCCCTCCAGTGGG	1452
Qy	33	-----LeuGluThrValThrValValGlyLysSerArgProArgAlaThr 47	
Db	1453	CGCCATATCAAGGATCTGGACAAAGTG--GTGGTCACCGCCAGCCCGTTCGCGATGCG	1509
Qy	48	SerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArg 67	
Db	1510	GCCGTGAGTTCAGCCGCCCGCGTG-----GAAAGTCTGGCGCGGAGCGTCTGGAC	1560
Qy	68	Gln---LysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86	
Db	1561	GAGTGGCAGTTCAGCTGGCGAAACCGTGGCCAGCTGCGCGGTACAGAGTTC	1620
Qy	87	GlnThrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLys 106	
Db	1621	AACCTTCGGCGCGCGTGGCGCGCGATCATCTCCGCGCCCTGGAGCGCGCGGTGGCA	1680
Qy	107	ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126	
Db	1681	GTGCTGGCGCGCGCTGTGACCCCAAGCGTGTCCACCGTCAGCAGGACCATTCGCCG	1740
Qy	127	MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeu 146	
Db	1741	GCAATCGAGCGCTTCCTGGCAACCAAGATCGAAGTGTCTGAAGGTGCTCAAGCGTGGT	1800
Qy	147	TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLys 166	
Db	1801	TACGGCTTCGGCGCATCGCGCGGTGTACACGTGTCTGACGCGCGCATCGCGAG--	1857
Qy	167	MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186	
Db	1858	ACCCCGGTGGACGGTTCAGCGCGCGCGCGAAGTACGCTTCAGCGCGCGGACAAAGAC	1917
Qy	187	LysLeuThrSer---GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThr 205	
Db	1918	GGCAACACCGACATGTTCCGCTGATCGCGCAATGGCAGCGCGCTGTCTGATCCAGCA	1977
Qy	206	GluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225	
Db	1978	GATGGCGTGTATCGCAACCAAGAACGACTACGACACCGCGAAGGGCGCGAGCTCAACAGC	2037
Qy	226	LeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArg 245	
Db	2038	TGGGTTCGATTC-----AAGTGGCTCGATCGCGTGTCTGCTCCGCGCAT	2085
Qy	246	LysArgPheTyrArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHis 265	
Db	2086	TGGGTTCGTTGTTGCGCATCGCGTTCGCGATTAACCTACGCGCAACCGCGCGAG	2145
Qy	266	SerHisGluTyrAspCysHisAlaSerIleIleThrGlnLysSerLeuIleAsnLys 285	
Db	2145	-----	2145
Qy	286	ArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnPro 305	
Db	2146	-----CCGGCGGACCCCTCCATCGCGAGCGC	2172
Qy	306	GlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAlaHisAsnGlyLys 325	
Db	2173	GGGGTGTG-----	2181

QY	326	ProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPhe	345
DB	2182	---CTGAAGTTGCAGCAGACCGCTACGAGCTCAAGGCGGCGGTGACCGATCCGTGG	2233
QY	346	ProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLys	365
DB	2236	GGCGAAGCAGCGCGCTGGTTACAGCTTCGCCATACCGATTACCGCATACCGAATTC	2295
QY	366	AlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeu	385
DB	2296	GAAAGCGCGCAAGTCCGACCGCTGTTCACCAAGCGGCGCAAGAGCGCGCTGGAAGCC	2355
QY	386	ArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLys	405
DB	2356	TCGTTTACGTTTGGTGGCGCTGACAGCTGCAATTCGGCTCGAG-----	2400
QY	406	SerSerAlaLeuSerAlaThrSerGluAlaVal---LysGlnProMetLeuLeuAspAsn	424
DB	2401	---GGCAGCGACAGCACCTTCAGGCGATCGCGCAAGAAATCCTTCGTGCCAAGACC	2454
QY	425	LysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTyrAspAsnPheThrLeu	444
DB	2455	GATACAGCTCGCTCGCGCATTCGCGTTCGACGCAATAGTTGGAGCGGTCAACGCC	2514
QY	445	GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle	464
DB	2515	GAAAGTGGCGCGCGCTGGACAAG-----GTCAAGTACCAG-----	2550
QY	465	AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln---Thr	483
DB	2551	---ACCGATATCGCGCTGCGTGGACCGCTGATTTCATCT	2580
QY	484	AlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeu	503
DB	2581	CCGACAGCTTTCGCGCAGCGCGTTCGCTTCAACGACAAATGGCGTCTGACCGCC	2640
QY	504	ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis	523
DB	2641	AACCTGGACACCGCGAGCGTGGCGCGCGAAGAAAGAACTGTTCGCGCAACGCCCGCAT	2700
QY	524	ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn	543
DB	2701	ATCGCACCTTCGCTTGAATTCGCGATCGCAATCTGAAGACCGAAGAGGCCAACCCAG	2760
QY	544	IleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArg	563
DB	2761	GCCGAGCTGGCTTGAATTTCCAGAACGAATGGGTGGATGCCCAAGGTGGCGGTACTAC	2820
QY	564	AsnArgPheGlyAsnTyrIleTyr-----AlaGlnThrLeuAsnAspGly	578
DB	2821	AACCGCTACACGACTTCGCTACATCGTCGACACCGCGCGCGCGTGTTCACGAA---	2877
QY	579	ArgGlyProLysSerIleGluAspSerSerGluMetLysLeuValArgTyrAsnGlnSer	598
DB	2878	---GAAGACAACGACTTCCTCGCATCGCGCATCGCGCGCGCGCGCGCGCGCGCG	2919
QY	599	GlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe-----LysProThr	614
DB	2920	GATCGCATCTTCACGGTTCGAAAGCGAGCGCACTTCACCTGCGCAACAGATAC	2979
QY	615	ProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuPro	634
DB	2980	GGTGGCTGGACCTCGCGGTTCGCGCATACGTCGCTGCGGTCTG-----	3027
QY	635	SerLeuProGlyArgGluAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAspGln	654
DB	3028	---AACGATGCGCGC	3039
QY	655	AsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp	674
DB	3040	AACCTCGCGGTATCCACCGGAGCGTTCGCTGCGCGCTGCGGTGGAT---GCCGAC	3096
QY	675	ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyr	694

QY 275 AspllelleTrrpGlnLysSerLeulleAsnLysArgTyrLeuGlnLeuTyrProHisLeu 294
 Db 1999 -----
 QY 295 LeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp 314
 Db 2000 -----
 QY 315 AspAspAlaHisAlaHisAsnGlyLysProTrrpIleAspLeuArgAsnLysArg 334
 Db 2009 GAGCGTGGTATCCTCCCGGTGACCGCGGTAGTTGGTTCGGATGTCATCAGGATCGT 2068
 QY 335 TyrGluLeuArgAlaGluTrrpLysGlnProPheProGlyPheGluAlaLeuArgValHis 354
 Db 2069 TATGACTTGAAGCTGGTTTAACTGATCATCGATCGCGGTGAGCGCGGTGCGCTGGCTTCAGC 2128
 QY 355 LeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePhe 374
 Db 2129 CTAGGTCATCTCAATATGATCATGATGTTGAGGGAACGAGGTGCGCACTACTTTC 2188
 QY 375 AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLys 394
 Db 2189 GGTAAAGCGTCTTCAGAGGGCGGAGTCGAGGCTCTTTCGATTTGGTGGTGGCTGGCGG 2248
 QY 395 GlySerTrpGlyVal -----GlnTyrLeuGlyGlnLysSerSer 407
 Db 2249 ACTGCAATTTGGTGCACAGGGGGCGACTCTTCTTCAGCGCTTGGCAGGAGTCAATTT 2308
 QY 408 AlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGln 427
 Db 2309 GTACCAAGACGAATAACAAATCAAT ----- 2335
 QY 428 HisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGlyGly 447
 Db 2336 -----GGTGTGTGGATGGCGCATACACTTTTGGGTGTTTCACGAGAAATTTGT 2389
 QY 448 ValArgValGlyLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspArgGlu 467
 Db 2390 GCGCGTGGTATCA -----GTTAAATACGATACGATAATGCGTACTCG 2437
 QY 468 AsnTyrTrrpLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPhe 487
 Db 2438 AATATATCAC -----CCAGGAGATTTG 2458
 QY 488 AlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis 507
 Db 2459 GCATTTAGTGTGATTTAGCACTCTTAAGCAATGGCGCTGACTTTGACGTGATCAT 2518
 QY 508 GlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn 527
 Db 2519 GCCGAGCGTGTCCGGTGGAGAGGAACCTGTTGCCAAGGTCGCGACATTCGCACTCTT 2578
 QY 528 ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAla 547
 Db 2579 GCTTATGAGTGGTGTGGCGGCTGAGAGAGAGAGGCTAACAGCTGAGTTGGG 2638
 QY 548 LeuGlyTyrGluGlyAspArgTrrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGly 567
 Db 2639 TTGCTATTCCGTAACGATGGAGTGATGTCGAAAGTTTCGACCTATTACAGTCGCTATGGC 2698
 QY 568 AsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAsp 587
 Db 2699 AATTCATTTAC -----TTAGTCGAT -----ACAGGGAGTACTTGGTTTGGGATGAT 2746
 QY 588 SerGluMetLysLeuVal -----ArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla 605
 Db 2747 GGACAGCGGTATTGGCGGTCCGTGAGTGGTGGCAAGCTAATGCAATTTTCAGGGAAT 2806
 QY 606 GluGlyGluIleTyrPhe -----LysProThrProArgTyrArgIleGlyVal 621
 Db 2807 GAAGGTGAGGCGACCTTTCAVTTGGCTAAATAACACGATGGATCTTGGGACCTACGTGTA 2866
 QY 622 SerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAsp 641

Db 2867 TTCGGTGACCGGTGAGTGGTTCGCTTGAATAAT ----- 2899
 QY 642 AlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
 Db 2900 -----GGAGGCAATCTGCCACGCATCGTCCCA 2926
 QY 662 AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAla ----- 678
 Db 2927 GCACGTTATGGGCGGAGCTACGCTGGAG -----GATGCTGGTGGCGC 2971
 QY 679 ---AsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArg 697
 Db 2972 ACATCGTTAGCGCTAGCGCTACGAAAGCAGATTAAGTACGGGTTAAGAGACCCCA 3031
 QY 698 ThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrTyrArgArgAsnThrArgTyrGly 717
 Db 3032 ACCGAGGATATACCATCGTAGTCGCAATTTGGCTTACCATTATGATGGTAGCATT 3091
 QY 718 GluTrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSer 737
 Db 3092 GCCTGGGAGGTCTTTTCGATGGTAACAATCTACCAACCGCGATGCTCGTGTTCATCT 3151
 QY 738 SerPheLeuSerAspThrProGlnMet ---GlyArgSerPheThrGlyValAsnVal 756
 Db 3152 TCATTCCTTAAGACGATGTCATGCTGGCTGGCGTAACTACTATCTGCTGTTTCGTATG 3211
 QY 757 LysPhe 758
 Db 3212 TTCTTT 3217
 RESULT 20
 LOCUS BX321862/c 295250 bp DNA linear BCT 23-APR-2003
 DEFINITION Nitrosomonas europaea ATCC 19718, complete genome; segment 7/10.
 ACCESSION BX321862 AL954747
 VERSION BX321862.1 GI:30180701
 KEYWORDS complete genome.
 SOURCE Nitrosomonas europaea ATCC 19718
 ORGANISM Nitrosomonas europaea ATCC 19718
 Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
 Nitrosomonadaceae; Nitrosomonas.
 REFERENCE 1
 AUTHORS Chain, F., Lamerdin, J., Larimer, F., Regala, W., Land, M., Hauser, L.,
 Hooper, A., Klotz, M., Norton, J., Sayavedra-Soto, L., Arciero, D.,
 Holmes, N., Whittaker, M. and Arp, D.
 TITLE Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and
 JOURNAL Obligat Chemolithoautotroph Nitrosomonas europaea
 PUBLISHED J. Bacteriol. 185 (9), 2759-2773 (2003)
 REFERENCE 2 (bases 1 to 295250)
 AUTHORS Larimer, F.
 TITLE Direct Submission
 JOURNAL Submitted (12-NOV-2002) Submitted on behalf of the Nitrosomonas
 Genome Consortium, the DOE Joint Genome Institute, Production
 Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
 USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
 larimerf@ornl.gov
 REMARK Nitrosomonas genome consortium
 FEATURES
 source
 1. 295250
 /organism="Nitrosomonas europaea ATCC 19718"
 /mol_type="genomic DNA"
 /strain="ATCC 19718"
 /db_xref="taxon:228410"
 complement (132..443)
 /locus_tag="NEI703"
 complement (132..443)
 /locus_tag="NEI703"
 /function="InterPro IPR003793 COGS COG1993"
 /codon_start=1
 /evidence=not_experimental
 gene
 CDS


```

/transl_table=11
/product="DUF190"
/protein_id="CAD85614.1"
/db_xref="GI:30180702"
/translation="MNGYQITFTTQDRHHAGKPLADWLMLAAEMGLGATLIPGSE
GMGDKRPHSVHFELSDQLLEVVMVVEEADQLFARLAEGVRLFYVKVAEFGIM
D"
gene
complement(454..837)
/locus_tag="NEI1704"
CDS
complement(454..837)
/locus_tag="NEI1704"
/function="InterPro IPR003691 COGs COG0239"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CRCB protein"
/protein_id="CAD85615.1"
/db_xref="GI:30180703"
/translation="MWKPIALALGSTLGGLLRWGLGLKLNLPDPVPPGTLVANLIA
GYVGVIAFAFHPNPSPEWRLVITFGCGGSTFTFSARIIVSLLRGLYAWAWSA
IAVHVAGSLIMTLAGIATVWFKSS"
misc_feature
order(complement(769..825), complement(664..729),
complement(568..633), complement(466..531))
/locus_tag="NEI1704"
/EC_number="4.2.1.1"
/notes="4 probable transmembrane helices predicted by
TMHMM2.0"
gene
1274..3895
/locus_tag="NEI1705"
CDS
1274..3895
/locus_tag="NEI1705"
/function="InterPro IPR00432:IPR001687:IPR002863 COGs
COG0249"
/evidence=not experimental
/transl_table=11
/product="mutS: DNA mismatch repair protein"
/protein_id="CAD85616.1"
/db_xref="GI:30180704"
/translation="MNKAQSSHTPMQOYLRIKAQHTDKLFLVRMGDFVELFYEDAE
KAAKLLDITLQRSGAGEPIKMAQVPPHAAQYLARILVRLGESIAICEQTGDPATSK
GPVERQVIRLTPGLTDAGLLEERSIVLALALHRSIGLAWLNLAAGDMRVLETS
SDNLTSELERLHPALIPESLDLPATLNNFAGPKRLPDMQDFVEHAMQOLTROFGTR
DLNARCCDLHNAIMAGALREYVRLTQQTATDGSSQGLPGHLHTLOVERQDAVLRMD
AATRNLEITLIRGEDAPTLLSSLDTCSTGSKRLRLHMLHRLNRILOQLRDYV
SLIGAQETTYAGIRQOKKHADIERITSLAKTARPKDLSGLRSLMKRLPGILSL
IATSAARVHRPIPPMQPDPPLLTQLLVRALQVPGAVIRGGVITADGFDLDELRLGL
QNCDEFTLQLEAREERTGIPNLKVEYNRVHGFYIEVTRAQGEKIIPDYRRRRLTKN
AERYTIPELQAFEHKTLASAREQALAREKMLYERLIEQLADFIIPLOIARSVAELDV
CAFARAALSGYTKPVTDDPVLIIIEAGRHVPVENQVEHYIANDVOLGAIPTRENQWML
VITGPMGKSTYMRQTMATVLLAHCGSFVPAQIARIGIPDQIFRIGAADLAGGRS
TWPMVMTAGLIRNATAGLSLVLDVEIGRGTSTFDGLALAPAIARHLLTONQSYLFA
THYFELTRIAEFPQAVNIHVAVEHKRRIIVFLHRIEGSPASRSYGLHVAALAGVPR
VIRNAAKILARDEQTLRSPOQTLFETVEENAKAVPASVHPVLIDYLERIHPDELTPR
GALQLYLKSLMNQTD"
gene
complement(3899..4378)
/locus_tag="NEI1706"
CDS
complement(3899..4378)
/locus_tag="NEI1706"
/EC_number="5.2.1.8"
/function="InterPro IPR001179 COGs COG1047"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="FKBP-type peptidyl-prolyl cis-trans isomerase
(PPIase)"
/protein_id="CAD85617.1"
/db_xref="GI:30180705"
/translation="MRLVKNVTVVSLLHYKMDQGNMVEETPEPVSLLHGGYDGIFFPAV
EEALHEKETGAALSQMEPDAFGEDYELVEENAKAVPASVHPVLIDYLERIHPDELTPR
FCIYTVREITDDVVVVDGNHFPYAGMTFTQFECTVTVRPAATAEELS GHVHGPHEH"
gene
complement(4393..5046)
/locus_tag="NEI1707"
CDS
complement(4393..5046)
/locus_tag="NEI1707"
/EC_number="3.1.26.4"
/function="InterPro IPR001352 COGs COG0164"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Ribonuclease HII and HIII"
/protein_id="CAD85618.1"
/db_xref="GI:30180706"
/translation="MAERRIPLKHEYAODGKVIYGVDEAGRGLAGPYAAACVWLDPA
DVIKGLADSKOLSEKKRISLADQIKORAWAIASAVEEIDRLNIIQASLLAMQRAV
VSLRPSNALVLDGNHAPRLDCEVQTVIRGDSLVAEISAASIIAKTARDIEMLRHE
AYPVYGPDRHKGYPYKAHLEAIRLGIITDIHRRSFAPCVGQSVSGARTTSTFNOKEA"
complement(5068..5529)
/locus_tag="NEI1708"
CDS
complement(5068..5529)
/locus_tag="NEI1708"
/EC_number="4.2.1.1"
/function="InterPro IPR001143 COGs COG0764"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Bacterial thioester dehydrase"
/protein_id="CAD85619.1"
/db_xref="GI:30180707"
/translation="MKQGRQTVMDIHEILKYLPHRYPLILVDRVSVLESKRHIAYK
NWSINEPYFSGHPPHVPMPGLIIIEALAAALLTIRSENMEKONGQVYVFGIDAV
RPKKPIAGDQVLKVEITRQKGIWKYKAAACAEVDGVVTEAQLMCTARAI"
complement(5546..6109)
/locus_tag="NEI1709"
CDS
complement(5546..6109)
/locus_tag="NEI1709"
/function="COGs COG2825"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative transmembrane protein"
/protein_id="CAD85620.1"
/db_xref="GI:30180708"
/translation="MRIPDDQYQGMFFRFVAVVVTMMFVLPVHSSAGEIKIGVNT
EKLVRSEPAIEAQKKIEREFQARDARIELSAQITALQOELEKNTGTVEEERLKE
RELAGRSQYQRAQQMRDLSLRQNEEYGLILERINQVIRELAEKQSYDLILQDLS
VYRSARIDITDQVIKVLNARESAARKP"
complement(6005..6070)
/locus_tag="NEI1709"
misc_feature
/notes="1 probable transmembrane helix predicted by
TMHMM2.0"
gene
complement(6136..8412)
/locus_tag="NEI1710"
CDS
complement(6136..8412)
/locus_tag="NEI1710"
/function="InterPro IPR000184 COGs COG0729"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Bacterial surface antigen (D15)"
/protein_id="CAD85621.1"
/db_xref="GI:30180709"
/translation="MKRLRLLIFFSLYSLGCMANDSLVVRDIRVEGIQTEAGTVESY
LPVKGVDLDSKXASAAIKALYATFFSDVKLSEGLLIIVQWERPAIAQISNGAK
EFDKDLKLEKQAGLSESRISFSLLEKAEKELKQYISRGKIAVKIITTTPLERN
RIGINFDIKEGKTARIKQINIVGNHVPEDDLVDLSLKTGPGMTWFTKDDQYSKQKL
SADLETIRSYLDRGYLENIESTQVITPDMKDIYITVNTTGGTPQYVSDIKLAGEL
LVPEEELRLIKLEPGGIFVREKLITSEIKLISDLRQNDGVAFANVNASPELDEKTKT
AFTFFIDPGRVYVRINISGNERTDEIVIRREFRQMEGWHSTEQINRSRQRVDRLD

```


QY	156	ValAspValAlaAspGlyValSerGlyLeu	175	Db	148013	TTGAACTGATCATGCGAGCGTCTCGGTGGAGGAGAACTGTTTGCACAGGCGCG	147954
Db	148823	GTCATATCGTTGATGAGCGATATGCTGAA	148767	QY	523	HisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAsnSerAsn	542
QY	176	LeuGlyLeuArgLeuSerGly	189	Db	147953	CACATTCGACTCTTGTATGAAGTTGGTCGCGCACTGAATAGGAGAGGCTTAC	147894
Db	148766	GCTGAATAGCTTTAGACGGTGGTGATAAACATGTAATACATGTTCCGCAATGAT	148707	QY	543	AsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyr	562
QY	190	SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyr	209	Db	147893	CAAGCTGAGTTGGGTTGGTATTCCGTAACGATTGGAGTATGCCAAAGTTTCGACCTAT	147834
Db	148706	GCTGGTAAC	148662	QY	563	ArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLys	582
QY	210	ArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSer	229	Db	147833	TACAGTCGCTATGCAATTCATTATAC	147786
Db	148661	CGTAACGAAAGGACTATCACACCCGAAAGGCGCGAGGTTAACTCGTTTCATCAGACACC	148602	QY	583	SerIleGluAspSerGluMetLysLeuVal	600
QY	230	ProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyr	249	Db	147785	TGGTTTGGGATGATGACAGCGTGTTCGCGGTCCGTCAGTGGTCGCAAGCTAATGCA	147726
Db	148601	AAATCAGCTCGATGGGTGCTTCGTCAGCGGTGATTTGGTTTCATT	148554	QY	601	AspPheTyrGlyAlaGluGlyGluIleTyrPhe	616
QY	250	ArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyr	269	Db	147725	ATTTTTCACGGAATTGAAGGTGAGCCACCTTTTCACTTGGCTAAATAATACGAGTGGGTCT	147666
Db	148553	GGCTGTGCGTCGCGGCTTTTCATGACAGCTATGGC	148518	QY	617	TyrArgIleGlyValSerGlyAspTyrValArgGlyValGlyAsnLeuProSerLeu	636
QY	270	AspAspCysHisAlaAspIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGln	289	Db	147665	TGGACCTACGTGTGTGGTGACGGGTTCGTGCTTGTGAAA	147621
Db	148518		148518	QY	637	ProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAla	656
QY	290	LeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys	309	Db	147620		147606
Db	148517		148509	QY	657	ProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIle	676
QY	310	GlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAsp	329	Db	147605	CCAGCATCGTTCAGCACCTTACGGGCGGAGCTACGTTGGGAG	147561
Db	148508		148464	QY	677	Asp	692
QY	330	LeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheGlyPheGlu	349	Db	147560	GACGTTGGTGGCTGCTACATCGTTGGTGTAAAGCGCTACGAGAGACAGACAGAGTAGCG	147501
Db	148463	TTGCATCAGGATCGTTATGATTTGAAGCTGGTTAACTGATCCATGGGTGAGCGGAGT	148404	QY	693	ArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArg	712
QY	350	AlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAla	369	Db	147500	GTCACAGAGACTCCACACACAGGATATACCATGTAGATGCGCATTTGGCTACCATATT	147441
Db	148403	GCCTGCTTTTCAGCTAGCTACACTCATATATGATCATATTGATTTGAGGGAACAGAG	148344	QY	713	AsnThrArgTyrGlyGluTyrAsnTrpValLysAlaAspAsnLeuLeuAsnGlnSer	732
QY	370	ValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro	389	Db	147440	GATGGGTAACATGCTGGAGGCTCTTTTCATGTAACATCTCCACCACTGCGAT	147381
Db	148343	GTCGGCACTACTTCGTAGCGTGTGTGTCAGAGGCGGAGTCCAGGCTCTTTCGATTT	148284	QY	733	ValTyrAlaHisSerSerPheLeuSerAspThrProGlnMet	751
QY	390	IleGlyArgLeuLysGlySerTrpGlyVal	402	Db	147380	GCTGCTGTTCATCTCTTCTTAAAGACGATGTCATGTGTGGTGGTGGTAACTACT	147321
Db	148283	GGTGGTGGTGGCGGAGCTGGTGTGTGTCAGAGGAGGAGCTCTACTTTCCAGGCGCTT	148224	QY	752	GlyGlyValAsnValLysPhe	758
QY	403	GlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeu	422	Db	147320	GTTGGTTGGTATGTTCTTT	147300
Db	148223	GCTGAGGAGTCATTTGTACCAAGACCAATAACATCAAT	148182	RESULT 22			
QY	423	AspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAspPhe	442	AE004513	15470 bp	DNA	linear
Db	148181		148143	AE004513/c			BCT 19-FEB-2003
QY	443	ThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAla	462	LOCUS			
Db	148142	CAGTAGATTTGGTGGCGGTGGTGAACA	148095	DEFINITION			Pseudomonas aeruginosa PA01, section 74 of 529 of the complete genome.
QY	463	LeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln	482	ACCESSION			AE004513 AE004091
Db	148094	AATGGCGTACTCGGAATTATCAC	148071	VERSION			AE004513.1 GI:9946669
QY	483	ThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer	502	KEYWORDS			Pseudomonas aeruginosa PA01
Db	148070	CCAGGAAGTTTTCATTTAGTTGGTGAATAGCACTCTCTAAGCAATGCGCGCTGACT	148014	SOURCE			Pseudomonas aeruginosa PA01
QY	503	LeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLys	522	ORGANISM			Pseudomonadaceae; Pseudomonas.
				REFERENCE			1 (bases 1 to 15470)
				AUTHORS			Stover, C.K., Pham, X.-Q., Erwin, A.L., Mizoguchi, S.D., Warriner, P.,
							Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
							Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,
							Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
							Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,

Wong, G.K.-S., Wu, Z. and Paulsen, I.T.
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
Nature 406 (6799), 595-964 (2000)
20437337
10984043
2 (bases 1 to 15470)
Stover, C.K., Pham, X.-O.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.D., Coulter, S.N., Folger, K.R., Kas, A., Iarbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 15470)
Pseudomonas aeruginosa Community Annotation Project (PseudocAP)
Direct Submission
Submitted (04-FEB-2003) Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada

This represents the February 3, 2003 version of the continually updated, reviewed, *Pseudomonas aeruginosa* PA01 genome annotation, from PseudocAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudocAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.
Class 1: Function experimentally demonstrated in *P. aeruginosa*.
Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.
Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

Location/Qualifiers
1. .15470
/organism="Pseudomonas aeruginosa PA01"
/mol_type="genomic DNA"
/strain="PA01"
/db_xref="taxon:208964"
126. .878
/gene="pruR"
/locus_tag="PA0780"
126. .878
/gene="pruR"
/locus_tag="PA0780"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in *P. aeruginosa*); Subcellular localization: Cytoplasmic; Subcellular localization confidence: Class 2; Alt protein name: probable transcriptional regulator"
/codon_start=1
/transl_table=11
/product="proline utilization regulator"
/protein_id="AAG04169.1"
/db_xref="GI:9946670"
MRACLVEAGHVFAGVDNRMILLIDGENDPTDPDOLLARLFEAPRYPTLADDFON
LLSVAGELARVSDLLTRALGGVLRALHRLFGFQPLASGPDLLERLDAHVR
NIARLITVAELAOCVSPSHFAQFKDSVGLTPTHTYLLKVRLLDCAARLLRESRLPLV
RIARECGSSASATTAMRRYLIGTIPKSLRG"

complement (999. .3062)
/locus_tag="PA0781"
complement (999. .3062)
/locus_tag="PA0781"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences); Subcellular localization: Outer membrane protein; Subcellular localization confidence: Class 2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04170.1"
/db_xref="GI:9946671"
/translation="MSSGLFPPSRPLWPLTLALACLIVSGETLGGADGRSELPSQVITANPLGNESPATPSVLEGBDELTLROKSLGETLNGLPVSVTFPGCASRPVIRGMDGDRFLNGVGDALDASSLSYDHAHPEDPNSVERLEVVRGPAALLYGNAIGGVNSFDNRIPSEFVDGIHSGELRYGCATDTRSRGALAGDGNFALHYDAASRENDVRIPGVAHSRQRIQIDGTGKHVRQNSDQGGAVGSGYHHEHAGLSIGTIDSNYSYPAE
DDVRLKMQDDRYAFASERIDLEGPFTSLKDAAYTKYHEKEIDEGTITFKNGYEGRIEARHRLPLGVLNGVGAQFANRSFALGEEAFVPHTEFSAALFALEEWKLSDRDL
SFGARLEHTRVDPDAKGNREFAENDQSFTTGLSTGAVYKLTPIWLSAATLSYTER
APTVELYANGPHATGVEYGDADADKEKAVSTDLALRFDNGVHKSGVGVYSPFSN
YIGLLASGRHNEGEVVAAGDDELPEYLYSGVRAEDFYGVEAQDRIHLLSPSFNFD
LELSGDYTRAKNKOTGEPLPRIAPLRLNTALILWELQOQARVDVDEHAASQHRVPEEL
STDGYTTIGASLGYNFDIGESRWLAFVKGTLNQTIVRYASSILKDRVPAAGRGIEAG
VKVAF"
3526. .6708
/gene="putA"
/locus_tag="PA0782"
/note="synonyms: prUA, prUB"
3526. .6708
/gene="putA"
/locus_tag="PA0782"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in *P. aeruginosa*)"
/codon_start=1
/transl_table=11
/product="proline dehydrogenase PutA"
/protein_id="AAG04171.1"
/db_xref="GI:9946672"
/translation="MFKASHVLQDQTRSAEFPFVTSANYAVDEEAYLGLLQALD
PGPDGIEAIRNARALIEFTRSDNAVDTLDALFQYSLDTQEGMLMCLAEALLRVP
DAATDALIRDKLSAADQORHLGQSDNVLVPAWGLVMTKVNIDERTGDRAPSVI
SLVORSGEPVIRAAWQKMLGKQFVIGRITISALXNKRPCRQGTYYTFDMLGEA
AUTADAKIMADYIQADITVGAEPFQVGPSPISIKLSALHPRIEVAQRVUSGEL
FANVLELAVRAKLVNGITIDAEADRLLESLYEKLMRDPAIAGWGEFGLVIOAYS
KRCPLVYMLTLGLRELGARIPRLVKGAYWDSIKQVQGDADGYPVYTRKEGDTTS
YLACARYLILSEHTRGVYVQFASHNAHTVAILALDARAAAGSERDFEFORLHGMG
DALYDTVTEKYNRVRIVAPGAHKDILLPYLVRLLENGANSFVHKLVDPVPVETL
TOHPVTQROFCTLANDRIPLPOAIPGAGKNSQINNNIONWNELELAYKPHLTRQ
WQAAPIVNGEKLSPAEHVRCPYDLERLVGSVQYASAGQAKALDVLQAWPRNATP
VDERAGILERLADLLEAGORLMALCTLEAGSLQDQIDDEIHEAVDFCRYAQARQR
LVREELRGPTGERNELFVEGRGIFACVSPNFPFLAIFLGISAALVAGNTVLAKPAEQ
TSLIAARTVELMFEAGLPKDVIALPGDGTALGGVFCRDARVAGVATGSTDRTAIRIN
ROLAEKPGAIALTIAETGQONAMIVDSTALPEQVTKDAVQSAFTSAGORCSALRVYV
QQDIAREVILLKGWAKLVKGVTEVRVSDPVDAEKAGLEOHIALKAKGLIA
EKVPGNLNGHFVAVPAVEISGIDELKENFGLVHVRYAAQDLKVVQAINATGFG
LTMVHSRNEETARRIEELARVNGNLYINRNIQIGAVGVQVPPFGHGSLSGTGPKAGGPNY
LIRFVSRSTSYNTTAVGNSALSLADAE"
7020. .8540
/gene="putP"
/locus_tag="PA0783"
7020. .8540
/gene="putP"
/locus_tag="PA0783"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in *P. aeruginosa*); Subcellular localization: Inner membrane protein; Subcellular localization confidence: Class 2"
/codon_start=1
/transl_table=11
/product="sodium/proline symporter PutP"


```

|||||
1811 GGTTCGAGAGCTTACACCGCGAGCTGTCCACCGCGCGGGTGTACAGCTGACGCGG 1752
Qy GluHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeu 517
Db ATCTGGTGGTGGCGCGCCCTCAGCTACACCGAGCGCGCGCCGCTTCTTACGAGCTG 1692
Qy TyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsn 537
Db TAGCCACAGCTCCGCGCGCGCCACCGGTACCTACGAGGTAGCGGATCGCGCGCGAC 1632
Qy LysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyr 557
Db AAGGAAAGAGGGGTCTCCACCGAGCTCGCCCTCGGCTTCGACACCGCGGTGACACGGC 1572
Qy AsnLeuAlaLeuTyrArgAsnArgPheGlyAsnThrValLeuTyrAlaGlnThrLeuAsnAsp 577
Db AGCTGCGGGTGTCTACAGCGGCTTCTCCAACTACATC-----GGCTGCTCGCCAGC 1518
Qy GlyArgGlyProLys-----SerIleGluAspAspSerGluMetLys 591
Db GTCTGCCATCGCAACGAGGAGGCGAAGTGTGCGCGCGCGGATGACGAGCGCTGCGG 1458
Qy LeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGlu----- 608
Db GAATACCTCTACAGCGGGTTCGCGCGGACTTCTACGCGTTCGAGCGCGGACCGCATC 1398
Qy ---lletyRPhLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArg 627
Db CACCTGCTGGAAAGCGCTACGGCACTTCGACCTGGAATCTCTCGGGGACTACACCGCG 1338
Qy GlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgPro 647
Db GCCAAGAACAGGAC----- 1323
Qy PheIleAlaGlnAspAspGlnAlaProArgValProAlaAlaArgLeuGlyPheHis 667
Db ACCGCGCAACCGCTGCGCGGATGCGCGCGCTGCGCTGGAACACCGCGG 1275
Qy LeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAla 687
Db CTGATCTGGAGATTG---CAGCAGTGGCGGCGCGGCTCGACGTCGAACACCGCGCTCG 1218
Qy GlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGly 707
Db CAGCACCGGCTCCGCGAGGAGAACTCTCCACCGGAGCTAC-----ACCACCTCGCG 1164
Qy AlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrp-----TyrValLysAla 725
Db GCCAGCTCGGCTACAACTTCGACCTCGCGGAGCGCGCTGCGCTTCTGTCAGGGC 1104
Qy AspAsnLeuLysGlnSerValTyrAlaHisSerSerPheLeuSerAsp---ThrPro 744
Db ACCACCTTACCAACAGCGCTGCGTACGCCATTTGATCTCTCGCGGACCGGGTGGCG 1044
Qy GlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db GCGGCGGAGCGCGCATCGAGCGGGGTGAAGTGGCGGCTC 1002

RESULT 23
AE016779/c 301995 bp DNA linear BCT 12-DEC-2002
LOCUS Pseudomonas putida KT2440 section 6 of 21 of the complete genome.
DEFINITION AE016779 AE015451
ACCESSION AE016779.1 GI:26557023
VERSION
KEYWORDS
SOURCE Pseudomonas putida KT2440
ORGANISM Pseudomonas putida KT2440
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 301995)
AUTHORS Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D., Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L.,

```

```

Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R.,
Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P.,
Holtzapple, E., Scanlan, D., Tran, K., Moazzaz, A., Utterback, T.,
Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J.,
Hohseil, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J., Timmis, K.,
Duesterhoft, A., Tummier, B. and Fraser, C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
2 (bases 1 to 301995)
Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D.,
Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L.,
Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R.,
Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P.,
Holtzapple, E., Scanlan, D., Tran, K., Moazzaz, A., Utterback, T.,
Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J.,
Hohseil, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J., Timmis, K.,
Duesterhoft, A., Tummier, B. and Fraser, C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
1..301995
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
89..712
/gene="sspA"
/locus tag="PP1320"
89..712
/gene="sspA"
/locus tag="PP1320"
/notes="similar to GP:9864180; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="stringent starvation protein A"
/protein_id="AAN66944.1"
/db_xref="GI:24982777"
/transl_table=11
/translation="MGATNRIACYPADHYSHRVLVLAEGSVSLIDVDPAHLPK
KLAENVPGSVFTLVDRDLALYESTVMEYLEERYPPMPVPYVARGNSLILMHRI
ORDWCALATVLDPSSSEARTEARKALRESLTGVSPLFSEFACFMSDQSLVDCCLL
PILWRLPVLGIELPRQAKPLLDYMERQFAREFPQASLSVEREMKLI"
693..1154
/gene="sspB"
/locus tag="PP1321"
693..1154
/gene="sspB"
/locus tag="PP1321"
/notes="similar to GB:X60221, SP:P24539, and
PID:509291; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="stringent starvation protein B"
/protein_id="AAN66945.1"
/db_xref="GI:24982778"
/transl_table=11
/translation="MKCAFKEPLMNSRPYLVALYEMIVDNDCTPHMLVNAEPKV
QVPDGASDGQIVLNTSPAVRSLLHNDVAFSEGRPSGVAHSLFVPGVAILGIARE
NGGMVFELEPSSNDGDELDVDDVDDGPPDDGGQPPRPTGRPSLKVVK"
complement (1216..1794)
/locus tag="PP1322"
complement (1216..1794)
/locus tag="PP1322"
/codon_start=1
/transl_table=11
/product="lipoprotein, putative"
/protein_id="AAN66946.1"
/db_xref="GI:24982779"
/transl_table=11
/translation="MTPKMLGMLALTLCLSVTCSSVLTSRNSPTEDRGRTRTIGSK
IDPSLLETAAVNIKANPNLDKGSIVVSYNGIVLAGOTPRADLKSLASQTASQV
QRVYKVNELQWQPSILARNDAWLTTKTKITQMLTDSAVPSSRIKVTENGIVYLL
GLVTRQANSATAVQGVGVQKIVKLFYIID"

```


complement (1791. .2378)
 /gene="gmbA"
 /locus tag="PP1323"
 complement (1791. .2378)
 /gene="gmbA"
 /locus tag="PP1323"
 /note="identified by match to PFAM protein family HMM
 PF01380"
 /codon_start=1
 /transl_table=11
 /product="phosphoheptose isomerase"
 /protein id="AAN66947.1"
 /db_xref="GI:24982780"
 /translation="WQSRIRRLPEASTTQQAMDILAPHEQASLVYVNALINEGKM
 LACNGSAGDAQHFSSSELNRRFRERPSIPATAITDSTLSIANDSYNEVFQSK
 IRALGQPDVILLATSGNSANVITQIAOAHDRMIVALTGDDGGMSALLLPDVE
 IRVPTVTARTQEVHLLAIHCLDLIDSLQFGSEE"
 complement (2471. .2845)
 /locus tag="PP1324"
 complement (2471. .2845)
 /locus tag="PP1324"
 /note="identified by match to TIGR protein family HMM
 TIGR00252"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein TIGR00252"
 /protein id="AAN66948.1"
 /db_xref="GI:24982782"
 /translation="MMAAASPRAGQAAETQALEYLOGQGLQILLARNWRCKGSGELDI
 MIDATYRVFVRVLRHGFGLGALDSIDGRKQRKLVIATILFQKEAHWGNYPGRFDV
 VALQGSHHVAGPRQLWLNKNAFEC"
 complement (2842. .4659)
 /locus tag="PP1325"
 complement (2842. .4659)
 /locus tag="PP1325"
 /codon_start=1
 /transl_table=11
 /product="lipoprotein, putative"
 /protein id="AAN66949.1"
 /db_xref="GI:24982783"
 /translation="MIACRLTALCALAALCAASSPSSLGELPRTPDASTIOLLE
 KAASKSREDALLRLSADLIYKQDFPPAARILEQVPLDTLKQAQVFASTLAEL
 AMSNQPKAATLAHPSLQRLASLPADLQARTYSVHAALEADGQAAQAQRILLAL
 PLLGQAATNDNDIALWLVSLPAEQIQQPTNDTLAGTSLAPAVKASGTLTQQQQA
 IEAWYQHPDPAQQLFLATLKLSELADQPLKTLIALLLQEGFLQAVARALDRGMA
 AHFQAQSGPAPAVQVPSDFISGLDIFYRQAQAGVQLVIGFLKPLVYQLAAKQP
 LPIITLALYADAGAQGPQDFQGLAAEDAREVARRADGMVRALVAVPSGEGW
 RVLAARPDWEGNGTLLAARIIAPVALAQIADLPQLQSEGRKASLGSTVGGSTA
 AQPSPQIDIFLFTSLTQQAQIKPTLNFQYAGDVPYATSNLYSAGSDINYNDM
 GTRCETFWLIDTSNLQVQVQVQWQPAQAGSLGRLYANGVDAYS LAPRLQLKALPDN
 RVLGUSGLSINANQVRERQDLPWAEFGGQVKRLPDTVR"
 4837. .5721
 /locus tag="PP1326"
 4837. .5721
 /locus tag="PP1326"
 /note="identified by match to PFAM protein family HMM
 PF00590"
 /codon_start=1
 /transl_table=11
 /product="tetraapyrrole methylase family protein"
 /protein id="AAN66950.1"
 /db_xref="GI:24982784"
 /translation="WCATVDVAGASKSTVGTLYVYVATPIGNLDDMSARALKVLAADV
 IAAEDTRSHIRLLQHFIDTFLAACHEHNEDEGRFLSKLLAGENVALVSDAGTELI
 SDGFHVLVRQAAGVSVVPFGCALIAALSAGLPSDRPIFEGLPAPQAQGRRLI
 EQVREPTLPIFYEAHRIILCLDEMAVFGGDRFALLARELTFTFTLKLFLAELR
 ARVAGDSQQRCECVLVGVGWSAPEGEQALSAEQRVLDLILABELPLKRAAALAAET
 GYRKLLYQLALEKOKAS"
 complement (5744. .6164)
 /locus tag="PP1327"
 /note="this region contains a gene with one or more
 premature stops or framehifts, and is not the result of a
 sequencing artifact; conserved hypothetical protein,

```

105 IleValValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
143873 ATTCGATCTCGGCAATGGCGCGTGCCTGGATGCTCGCTGCTCTACGATCAT 143814
125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleValGluArgGlyProValThr 144
143813 GCGGTGCGGTGAGCCCGGTGACCTGACGCGGTGCAATGCTCGTGGCCCGCGCGCC 143754
145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysPro 164
143753 CTGCTTTACGGCGCAACGCCATTTGGCGGGTGTCAACACCTTTCGACACCGCATCCG 143694
165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyValGluArgLeuSerSerGlyAsn 184
143693 GACTCG---CCGATCGAGGCGCATCCCGGTGCGGTGAACTGCTTACGCGCGCGCGGAC 143637
185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
143636 ACCACTCGCAGCAGCGCGGCAAGCTGGAGCGCGCGCGCGCC---TTGCGCTGCGAC 143580
205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys 224
143579 CTGGATGCCAACAGCCCGCATTTCAACGACCTCGCATCCCGGCTATGACGCGCGCC 143520
225 ArgLeu-----ProAspSerProArgPheAlaAsnGlyGlnHisArg 239
143519 AAGTGGCGCATGCCGAGCGCCCGCGCAGCAACACCGCTCGAAAC-----143472
240 AlaValLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGln 259
143471 -----AGCCAGCGTCCCGCAGGAC 143454
260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
143453 GCGGCGCAATCGGTGTGCTTACCCTGGAGCACCGCTACCGCGCTGCTGCTATC---143397
280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
143396 -----AGCCGCTAC-----143388
300 ValAspTyrAspAsnProGlyLeuSerGlyPheHisAspAspAspAlaHisAla 319
143387 -----GACAGCAACTACCGCTCG 143370
320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
143369 GTGCCCGAGTCGGCGTGGC---CTGGACATGAACAGGATCACTACGCTTCGCTCC 143313
340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
143312 GAATCGCGCATCTCGAGCGCGCGTTCAGCTCGTCAAGGTAGATGCTGTTACACCGAC 143253
360 TyrHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnThrGln 379
143252 TACGAACAGCGAATCGAGGAGCGAGTTCATACCACTTCAAGAACAGGCGTAC 143193
380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
143192 GAAGCCCGCATCGAAGCCCGCTCACAGCCGCTCGGCGCGATTCAAGGGGATCGCTGCC 143133
400 GlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnPro 419
143132 CAGTTCAGCGCGCAACGAATCTTCGCGCTGGC-----GAGGAAGCC 143091
420 MetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlnGlnAlaAsnTrp 439
143090 TTGCTCCGCGCACCGATACCGACGCTCGCGTGTTCATGCTGGAGCAG-----TGG 143037
440 -----AspAsnPheThrLeuGluGlyGlyValArgValGlnLysGlnLysAlaSer 456
143036 CAAGCCACCGAAGCCCTGAACCTTGAAGCTGGCGCGCGCATGGAG-----CAC 142989
457 IleArgTyrAspLysAlaLeuIleAspArgLysGlnTyrTyrLysGlnProLeuProAsp 476

```

```

142988 ACCGCGTAGACCGCGAGCCCAAGGCAACGAGAACTTCGTCGAC-----142944
477 LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsn-----TrpTyr 494
142943 ---GCCGACGCGCGCAGCAGCTTCAACGCTTACGCTGTCTGGGGGTGTGTACGAG 142887
495 PheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThr 514
142886 CTCGACCCCAATCTGTGTGCTGGCGCCCAACCTCGCTACACCGAGCGCGCCCTACCTTT 142827
515 GlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys 534
142826 TACGAGCTGTACGCAATAGCGCGCAGCTAGCCACCGCGCGGTGTAAGTCGGTGAGCGCC 142767
535 HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArg 554
142766 AGCTGACACAGGAAGAGCCATTTCCGCGCAGCTGGCGCTTCGACACCGCAC 142707
555 TrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIle---TyrAlaGln 573
142706 CACAGGCGCATGTGGTGTGTTCTTACAGCACCTTTCGCAACTACATCGGCTGATGGC 142647
574 ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu-----589
142646 ACCGCAACACCGCGAGCGCAGCATGACCATGAGGAGGAGGATCAGATCATGATCAT 142587
590 -----MetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla 605
142586 GATCATGAGGCTTCCCGGAATACAGTACCAAGCGCTGCGCGCGCTTCTACGCGATC 142527
606 GluGlyGluIleTyrPheLysProThrProArg-----TyrArgIleGlyVal 621
142526 GAGGCGCGAGGACCGCTGGCAGTTGCCGAGAACCGTTACGCGCAGCTTCGCGCTGAGTTG 142467
622 SerGlyAspTyrValArgGlyArgLysAsnLeuProSerLeuProGlyArgGluAsp 641
142466 TCCGCTGACTATACCCGAGCC-----AAGAACCTCGACAGT-----142431
642 AlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
142430 -----GGCAGCCACCTGCCCGCATCGCCCGC 142404
662 AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
142403 CTGCGCTTGAACAGCGCGCTGTCTGGAGAGCTG---GACCGCTGCACGCGCGGTAGAC 142347
682 TyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHis 701
142346 GTGCAGCAGCTGCTGCTGCGCAGCACCGCAACCCGCAACAGCAACAGCAGCATGGCTAC 142287
702 HisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrp 721
142286 ACCACTTGGCGCGCAGGCTTGGCTTACCGTTTCGACATTTGGCCAGCAGCGTGGCTGGCG 142227
722 TyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSer 741
142226 TTGTGCGCGCGCAACCTGACCGCAGCGCTGCTATGCGAGTTGCGATTCGATCTCGCGC 142167
742 Asp---ThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
142166 GATATCGCACCGCGCGCGGCGGTAGCTGGAAGTTCGGGTTCGCTACCACTTC 142113

```

RESULT 24

AE005695

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

11626 bp DNA linear

BCT 12-JUN-2002

section 21 of 359 of the complete

genome.

AE005695 AE005673

GI:13421330

Caulobacter crescentus CB15

Caulobacter crescentus CB15


```

/note=identified by match to PFAM protein family HMM
PF00795"
/codon_start=1
/transl_table=11
/product="hydrolase, carbon-nitrogen family"
/protein_id="AAK22199.1"
/db_xref="GI:13421337"
/transl_table="MARTLSVAIAQTSYGMQLQANIKKTEGFIREAAKGAQVILPSE
LFQGPYFCVQEEBWFQAHPREHVPVKAIAPLAGLVVPIISIPERGGPHVFNLS
VMADAGSLMGVYKSHIPDGPYMEKYFPGDTGPKVWDTFRGIGVIGICWQWYVP
ECARALMGALFYTAIGSEPHDASLDLTPWRAMQGHAVSNVIPVIGANRGTG
EPWGYNGGQGTFTGSSFVADHRGDLVSELGRADEGLVSATFDLDFUTTHRAAWGFFR
DRPELYTALSSGRPA"
/gene=".9282"
/gene="CC0213"
/gene="CC0213"
/gene="CC0213"
/note=identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="acetyltransferase, GNAT family"
/protein_id="AAK22200.1"
/db_xref="GI:13421338"
/transl_table="MIETDLVLRPAVDADADATADLNARPEVGAWLGVRSAESDA
FVDRVMAHEAQNGFQFVVERKADVLVGTGTVWVPPAMNMDPTTEIGHKRFLEPVWG
QGYATERATALATGFDVLRPELIAFTARSNLASQAVMRIGMTAPAPDFLHPLRA
EGDPLRHVTPEAHPQAGRDGD"
/gene=".11473"
/gene="CC0214"
/gene="CC0214"
/gene="CC0214"
/note=identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAK22201.1"
/db_xref="GI:13421339"
/transl_table="MVSGPNMIRVVALSASLVLAATAAHAKDAAPASQADAEV
SKVIVTAAPYGVSDALATASVAVLDRTDLDAAPKGLDALAGLPGVRSFTFGAGAR
FVVRGLAGFVOVITVQGIIDASALPDHVAQTPDGEAREIEVLRGPAALAYGSAI
GAVNVIDDRISTQPDIGMSRLLAGRTGDDSYALSGAVHATVGMVLTDLAKRE
SKDYKIPYPSARQLALEGETAGGAGRLNSAVDLETFGAGLSYVVDGKGFVGMSTK
RTDSTYGVPHAHHEGEAEAGHEEASVTIGLKOTRIDLRGEVDALGRFPAKVRFSG
GHADYTHTEFGDVAQTKFTSDGVEGLELVQTERGGWKGAVGQALRRNFDAIGDRA
YVPTKITEFAGTQORLELDGYEGGLELDTRFLSLKGADEFNLSSVGAFWRP
TTSEFVGLALSRSRAPTESSELFAGPHATRGFEIGDADLKKEBTATLSUEATLHYGE

```

Alignment Scores:

```

Pred. No.: 1.13e-28 Length: 11626
Score: 636.00 Matches: 206
Percent Similarity: 40.23% Conservative: 107
Best Local Similarity: 26.48% Mismatches: 297
Query Match: 15.76% Indels: 168
DB: 1 Gaps: 23

```

US-09-936-377-2 (1-758) x AE005695 (1-11626)

```

QY 22 AlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLys 41
Db 9518 GCTTCGGCGGACAGCCGAGTCTCGAAGGTCTATCGTCCCGCGCACCTACGGGTC 9577
QY 42 SerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle 61
Db 9578 TCG-----GCTGATGCGCTCACCGCCAGCGCTCGCGGTGCTG 9613
QY 62 SerGlyAspThrLeuArgGlnLysAla---ValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db 9614 GATCGTACCGACTCGATCTGGCGGCCCCCAAGGGCTGGGCGACGCCCTCGCGCGTCTG 9673
QY 81 ProGlyIleHisAlaSerGlnTrpGlyGlyAlaSerAlaProValIleArgGlyGln 100
Db 9674 CCGGGCGGTGCTGACGACGCTTCGGCGGGCGGCCAGCGCGCGGTGCTGCGGGTCTG 9733
101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
9734 GCGGGCGCGGTGCTGCGAGTCTGACCAACGCGGTGGCGCAGATCGACCCAGCGCGCTG 9793
121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluLeuLeuArg 140
9794 TCGCCCGACGACGAGTCCGACCGATCCCGGAGCAAGCGGAGCGGATCGAAGTGTGCGC 9853
141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
9854 GGTCCGCGCGCTGCGCTATGGCGCTCGGCGATCGGCGGTGCTCAACATCATGAT 9913
161 GlyLysIleProGlyLysMetProGluAsnGlyValSerGlyGlyLeuGlyLeuArgLeu 180
9914 GATCGGATCTCCACCCAGCAGCGATCGACGCGCATGACGCGTGGCTGCTCGCTCGCGC 9973
181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyLeuAsnIleGlyLeuGlyLysAsn 200
9974 GGGACCGGTGACGACAGCTACGCGGTGTCGGCGCGCTCCACGCCACGCTCGGCGCG--- 10030
201 PheValLeuHisThrGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
10031 ATGTCCTCGAGCTCGACGCGCTGAAGCGCGAGAGCAAGACTACAAGATCCCGGTCTAT 10090
221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAla----- 234
10091 -----CCGGAGTCCGCCGCTGAGTGTGGCTGCTCGAGGCGCGAGACGGCC 10132
235 ---AsnGlyGlnHisArg-----AlaVal 241
10133 GAAGTGGCGCGGTGCTGCTTGAGAACAGCGCGTGCATCTGGAACGTTTCGCGCGAGGC 10192
242 LeuGlyTrp-----ArgLysArgPheTyrArgThrTyrSerAspArgArgAspGln 259
10193 CTCTCTATGTCGCGCACAGGCTTTGTGCGCATGTCGATCAAGCGCGACGACTCCACC 10252
260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
10253 TATGGCGTCCCGCGCACGCCCATGAG----- 10279
280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
10279 ----- 10279
300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
10280 -----CATGAGCGCGCGCGCGCGCGCGG 10303
320 HisAlaHisAsnGlyLysProTyrPheLeuAspLeuAsnLysArgTyrGluLeuArgAla 339
10304 CACGAGGAGAAAGTGGGTCCACCTCGGTCTGAGCAGCAGCGGTATTGACCTGCGTGGC 10363
340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
10364 GAGTACGACGCGCATCTGGGCGCTTCGCAAGGCTGCTTCGCGCGCGCGCGCGGAC 10423
360 TyrHisAspGlyLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGln 379
10424 TACCCACACCGAGTTTGGGTGACGCGTTCGCGCACCGCATCAAGTTCACCTCGAGTGGCTAT 10483
380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
10484 GAGGGCGCTCGAGCTGTTGAGCGGTTCAGCGCGGTGGCGGTGGAGGCGCTGGGCGCTC 10543
400 GlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnPro 419
10544 CAGGCTCTACCGCAACTTCGACGCGCATTCGCGCGAGCGGCTATGTGCGCAAG----- 10597
420 MetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp 439
10598 -----ACCAAGATCACCGAGTTCGGCGCGGTTCACCCAGCAACGCCCTCGAGCTT 10645
440 AspAsnPheThrLeuLeuGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyr 459

```

```

Db      10646 GACGGTTACGGTACGAGCGCGTCTCGGATCGATACCGCAACTCGATAGCTGAAG 10705
Qy      460 AspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAla 479
Db      10706 GCGAAGGCT-----GACTTCAACAAC----- 10726
Qy      480 HisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsn-----TrpTyrPhe 495
Db      10727 -----CTGTGGGCTCGGTGGCGCCCTTCTGGCGGCG 10759
Qy      496 ThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGln 515
Db      10760 ACCACGAGAGCTTCGTGGCGCTCGGCTCGCGCTCCAGCGCGCGCCACCGAGTCG 10819
Qy      516 GluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHis 535
Db      10820 GAGCTGTTCGCGAAGCGGCCACCGCGGACCGCGTGGTTCGAGATCGGCACCGCGAT 10879
Qy      536 LeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrp 555
Db      10880 CTGAGGAGGAGACCGCCACGCTCCCTGGAGGCGAGCTCAATTACGGCGCGAGCGGTT 10939
Qy      556 GlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeu 575
Db      10940 TCGCGGCGATCGCATCTATATCGCGCGCTACGACGGTTCATCGATCTGCGCCCGAC 10999
Qy      576 AsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr 595
Db      11000 GGTGAC-----GAGGAAGATGTT---CTGCGCTCTATCGCTAC 11035
Qy      596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluLeuTyrPheLysProThrPro 615
Db      11036 GTTCAGACCGATGCTGAATTCGCGGTTTCCGAGCGGATC----- 11077
Qy      616 ArgTyrArgIleGlyValSerGly-----AspTyr 625
Db      11078 GCCTATCGCTCGGACCGATGTCAGCGTCTGGTGAACCTGATCGCGGTGGGACTTC 11137
Qy      626 ValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsn 645
Db      11138 GTCCGGGCGC----- 11146
Qy      646 ArgProPheIleAlaGlnAspAspGlnAsnAlaPro---ArgValProAlaAlaArgLeu 664
Db      11147 -----TCGTCGACCTTGGACCGCCCGCGGATCCCGGATCCCGCTACTCGGTC 11191
Qy      665 GlyPheHisLeuLysAlaSerLeuThrAspArgIle---AspAlaAsnLeuAspTyrTyr 683
Db      11192 TCG-----CTCAAGCGCTCTTACCAAGCGCGATGTCGGCGCATGGAGGTGCGC 11245
Qy      684 ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMet 703
Db      11246 CGACCGGTGGCGAGGACCGCTCGCGCATGAGTCGCCACCGCGCTACACGGTG 11305
Qy      704 LeuAsnLeuGlyAlaAsnTyrArg-----ArgAsnThrArgTyrGlyGluTrpAsnTrp 721
Db      11306 CTCAACCGCCCTCGCGTGGAGCGCGCGCGTAAATCGCGCGTG-----CGCGTG 11356
Qy      722 TyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerPheLeuSer 741
Db      11357 TTCCTCATGCGCCCAACCTCGAAGATGAGGAGATGCGCGAGCATGCTCTCTCTCGAAG 11416
Qy      742 Asp---ThrProGlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
Db      11417 GACATCGCGCGTTCGCGGCGCGCAAGTTTCGCGCGGATCGCGCTTC 11470

RESULT 25
LOCUS   AP005943/c
DEFINITION Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 9/31.
ACCESSION AP005943 BA000040

```

```

AP005943.1 GI:27350476
Bradyrhizobium japonicum USDA 110
Bradyrhizobium japonicum USDA 110
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
1
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,
Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K.,
Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.
and Tabata, S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110
DNA Res. 9 (6), 189-197 (2002)
22484998
12597275
2
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,
Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K.,
Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.
and Tabata, S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
DNA Res. 9 (6), 225-256 (2002)
22485002
12597279
3 (bases 1 to 300400)
Kaneko, T.
Direct Submission
Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/rhizobase/,
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)
Location/Qualifiers
1..300400
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
48..1313
/gene="dada"
48..1313
/gene="dada"
/feature="ORF ID:bl2222"
/codon_start=1
/transl_table=1
/product="D-amino acid dehydrogenase, small subunit"
/protein_id="BAC47487.1"
/db_xref="GI:27350477"
/translation="MKVLILGSGVIGVTSAYVLGAGHDVTVVDRPEPALETSPANA
GEVSGSSPWAGFGVPVAVKWLMLKHPVIRPKLDVPMWLLKMLNCTARVA
VNSGRMPIASRDLRDLDIGIQYDERSQGLTLQFHQAQDGTGEDIAVLKQY
GVPEVLGREGCIAEPALAGVKEKFGAGLRLPDQDCHMTQALAKHAQALGVRF
MFTQIDRIYTDGAVSVTVTSAGVLAQVLAALGWSRLVAPLGLTSLPVPVKG
SITVPIKASGAPESVTMDESQYKVAITRLGNIRVGTABISGSSKLYDARRATLH
SLTDLFRGGDLTKATPWSGLRPMTPDGPVIGTQYVLAHLNHLTGHGTLCMTSCGSG
RVLDMLSGKXPEVDVVALTDYAHFRG"
complement (1411..2088)
/gene="b112223"
complement (1411..2088)
/gene="b112223"
/feature="ORF ID:bl12223"
/probable_transcriptional_regulator"
/codon_start=1
/transl_table=1
/protein_id="BAC47488.1"
/db_xref="GI:27350478"
/translation="MSGGKMRKPAAPKAKLAKTVPKAMTKPAEPAMDVAVGR
IRDRYRQFSLTVAARTGLSGLQIERGLSSPSLRALATLSVLGVMAALFGS
NPSADGSDVVTGELORPELKLWESGSKOLLSAAGSDNRLNLFVHIDPGSGAGDE
LYTHDGEAGLVLDGEMMLITVDSETWLSLTKGDSFRFASRPHRESNPAQAKAVLVW

```

```

NCVTGAG"
2252..3016
/gene="blr2224"
CDS
2252..3016
/note="blr2224"
/codon_start=1
/transl_table=11
/product="ABC transporter amino acid-binding protein"
/protein_id="BAC47489.1"
/db_xref="GI:27350479"
/translacion="MALKRFRVQAAVAALAIAGVVPASAQOVLKVGSTPTGIPFTFELDT
KNTIIGIMWDLITEIGKAGLVNQIEPMSFSALESLSKIDIIIAAMFITAPEKKE
VWFSPIIYTGSLVVPKNDTKAYATODDKGETGAQVGFADALKKSGLEADVK
AVDTIPILRDVNTGRLKAGYADYPILAVNLKOGFPEVRLVDSYKPVTVGSGVIGVR
KGPSALLGKINASLAKLAKNGTIDIKLEWQKRAQ"
3045..3698
/gene="blr2225"
3045..3698
/gene="blr2225"
/note="ORF ID: blr2225"
/codon_start=1
/transl_table=11
/product="amino acid ABC transporter permease protein"
/protein_id="BAC47490.1"
/db_xref="GI:27350480"
/translacion="MKGFWDAAEFPIILMSGVALTIVVTIGSLLSVLGLTWAMMR
VSGIKVLSMLASLIRIGIPITIVLLFYLIFVMPDLGVTLSALQAAIIGLGIAYSAY
QAFNFRAGIADIKQGTIEAAQSIGMGWMLTMRVLPQAVRIVLPYGNVMIMLKDS
SOASTIVASLALQGLIASSTFKNTSVFTLVALMTWSIPLILLVRHFEKRAKK"
3695..4426
/gene="blr2226"
3695..4426
/gene="blr2226"
/note="ORF ID: blr2226"
/codon_start=1
/transl_table=11
/product="amino acid ABC transporter ATP-binding protein"
/protein_id="BAC47491.1"
/db_xref="GI:27350481"
/translacion="MIELTDVHSFGKVEVLKIGITKVEKGVVVICVPGSGSKSIL
RCINGLESYGEINVEGLKVDKDPASIVNVRTQVSMVFORFNLPHRTVLENVVEGP
LYNKEARAQVLEGRGALLAQVGLAEKADAPPLQGGQOORVAIARALAMQKALIF
DEPTSAIDPELVGLVGRKLDADGTMVVTWEMGFARDVADRLVIDGGVIVVEQG
PAKSVLNGQHPRTQDFLRVHL"
4439..5725
/gene="blr2227"
4439..5725
/gene="blr2227"
/note="ORF ID: blr2227"
/probable oxidoreductase"
/codon_start=1
/transl_table=11
/protein_id="BAC47492.1"
/db_xref="GI:27350482"
/translacion="MTARLPLPSSIYADTAVATPPLDADKTVSAIIGGTYGLS
TALHARQGVLEALVEAQEWGASGNGGHTNPLKHPDQIEADFGTGLGRMTLEF
SYGNPTHELIRYQIIPCARQNGTLRAAYNEASAAAEKTAQOCIRRGMPVYTLNR
EOLREMTGTDYICAMLDARGDDLPLSYARGLARAISAGAKVHGETPALSLRPGA
HWLETPRAVRAEKVLLATNGTDDLPALRSTIVPVTSSIVASAPLSAEVARSIMP
TRPVYSGHTIVYRIDQNRLLMGGRGPMWISSPESDVAYLIRYAEIRLWPLQKGA
WTHGWNRLAITKDYHPHVEPAESILISLGCNGRGVALSTWGAQLARLLIGAKAE
IDMPTVTKIPFIMHAFWPGVVTAVITGRVDRRLGI"
/gene="blr2228"
36145..6145
/complement (5777..6145)
/gene="blr2228"
36145..6145
/complement (5777..6145)
/note="ORF ID: blr2228"
/unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC47493.1"
/db_xref="GI:27350483"

```

```

/translacion="MCQHALSRCAIHLRBRHDSSTFAGNLAHSPFIDRTSALAPLE
NMKRLAASLIIAAGFTLTNPAPAFKGGHGHGHHYDGTGPPQWSHG
RKVCWRGHCPCPPGLWKQRC"
/complement (6201..7916)
/gene="blr2229"
37916..6201
/complement (6201..7916)
/gene="blr2229"
/note="ORF ID: blr2229"
similar to "ACY1 bifunctional acylase"
/codon_start=1
/transl_table=11
/protein_id="BAC47494.1"
/db_xref="GI:27350484"
/translacion="MFSLAAMCHCPRASPNALFKSPSQVHRSQDORRQTDRIIMP
HOFSLAQTVRPVAVTKGIVAAQSRRAAEVGAQVLAAGGCDVDAIVATTMALNVLEP
WNSGTGGGAMVLYARENTEVIDYGCAPQSLRVADYPLSGGGAADLPFPVRVKD
DRNTHGCAIAPVGVAGMEARHAKLPKDIAPAALAGSLADMMWTTLTIAS
SAADLRYPASAAFLKGLPFPAPWGIKSETRLSQKLKATLSHLAEAGRPDYQGD
LAKSIADIKADGSSLSVEDIAAFRAHLREPLAIPYRGKVFATPELTAGPTLAHLR
LLQNLKPAGAPDAATVAEYALALQAAAYRERLKDMDGDKRSIAGAYLAPACTHFS
VVDHGNVAAVTQTLSSFGSKYVTPHTGIAMNNGIMWFDPGTGTTSLAFKRCCLTN
YTPVIAETRDGRLAIGASGRRILPAVQLASFAMDGMDLDAIHQPRIDASEGAV
VIGDVLPAIDTRKALAAAFDEESRVQALPQKFACPSPVMREGEVNFQVAVEIFQPWD
AAEA"
7934..8707
/gene="blr2230"
8707..7934
/gene="blr2230"
/note="ORF ID: blr2230"
/codon_start=1
/transl_table=11
/product="enoyl CoA hydratase"
/protein_id="BAC47495.1"
/db_xref="GI:27350485"
/translacion="MAGVKQARDGAVGILITDEPASLNAMTPDLIGALAAAVGGMQD
EGVRALITGARGPCSONLKASEALGEDIAAGVMRPFYPAFKALRECRYPVVVAVN
GVAAGGSLAMAGDIIVAAASAFIQVFSIALVDPDLSGTWLLPRLNGRORALEML
LINEPTARAOEIGLIRVQVDDKLMGEALVLRLEAGPTALVATRALVESEHAT
YEAQFRRIELQMKIRKSADALEGRDAFVEKKAKFTGR"
8664..9227
/gene="blr2231"
9227..8664
/gene="blr2231"
/note="ORF ID: blr2231"
/unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC47496.1"
/db_xref="GI:27350486"
/translacion="MPSSRSARRSPGGEPEKTSMPGTCPGIASLENRFAASGETTAS"

```

Alignment Scores:

Pred. No.:	9,09e-17	Length:	300400
Score:	469.00	Matches:	200
Percent Similarity:	38.48%	Conservative:	124
Best Local Similarity:	23.75%	Mismatches:	335
Query Match:	11.62%	Indels:	184
DB:	1	Gaps:	34

US-09-936-377-2 (1-758) x AP005943 (1-300400)

```

Qy 4 ThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGln 23
Db 283418 ACGGCGCGAGCCGATGTCGCGAGGTGGTACCGACCGCGCCGTCCTCA-TGTCGC 283360
Qy 24 AlaHisGluThr-----Glu-GlnSerValGlyLeuGluThrValThrValGlyIly 41
Db 283359 GCGCACCGGACGCGTCCGCGAGCCAGCAACACGCGCGCGAGCTACGCGCGCGCG 283300
Qy 41 sSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThr-----AlaSerAs 58
Db 283299 GTCTGCCCGCGCGCTCAGCAGGCGGTGCTGCCCATCTGTGACCAACACGAGTTCGCCCGCT 283240

```

```

Qy 58 p1y5lleleSerGlyAspThrLeuArgGlnLys--AlaValAsnLeuGlyAspAlaLe 77
Db 283239 CACGTTGGTCCGACAGAGAGATCCCGCGAGGGCGGGTCCAGTCCGGCATCTCTT 283180
Qy 77 uAspGlyValProGlyLeHisAlaSerGlnTyrGlyGlyAlaSerAla---ProVa 96
Db 283179 GTTCTCCAAGCGCGGCATCACCGCTCGAGCTTCGCGCGCGCGCTCCAGCGCGCGAT 283120
Qy 96 l1leArgGlyGlnThrGlyArgGlylleValLeuAsnHisHisGlyGlnThrGlyAs 116
Db 283119 CATCCGGGGTCCAGCTCAACCGCTCGCATCGTCCGAGACCGCACCAATTCGGGTGG 283060
Qy 116 pMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnInVa 136
Db 283059 CGCTCCAGCTCGCGAGGATCACTTCGTCCTCCGATCGCTCGCGACCAACACAGGT 283000
Qy 136 lGluLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVa 156
Db 282999 CGAGTGTGCGCGCGCGGTGCGCTCCGTCAGCTCGACCTCGATCGCGCGGTGCT 282940
Qy 156 lAspValAlaAspGlyTysIleProGluLysMetProGluAsnGlyValSerGlyGlu-- 175
Db 282939 CAGCGCGACCAACACCGGATTCGGATCGCTCGCGAGTTCGCTCGCTCCAGAC 282880
Qy 176 -LeuGlyLeu----- 178
Db 282879 CTACGGCTCCCGACCAAGCGCGCTCGGCAACCGCGAGACCTCGCTCGCTCACGGC 282820
Qy 179 -----ArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGl 196
Db 282819 CGAGACGCGCTCCGCAATTCAGCTCGCTCGATCGCGCGCTGAAAGCGGTGCTCGA 282760
Qy 196 yLeu---GlyLysAnpPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTy 215
Db 282759 CACCGCGCGCGCAATTTGCTTCATCGCGAGCTCATGCGCGCACCAACCGGACTA 282700
Qy 215 rAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgPheAlaAs 235
Db 282699 CGGCATCCGAGCTATCCTATCTGAAC-----GACCAGACGGATCCGTCGCCAA 282649
Qy 235 nGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAs 255
Db 282648 CGGC-----CGCCAGCGCAACTCGGC 282628
Qy 255 pArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlnTyrAspCysHisAlaAs 275
Db 282627 GACGCGCTCGACGCGGCTCGATCGCGGCTCTCTAC----- 282591
Qy 275 p1leleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLe 295
Db 282590 ----TTCCTCCAGGGCGGTATATCGGGCT-----GCGAT 282559
Qy 295 uThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAs 315
Db 282558 CACGAGACGACTCGCTCTACCATCCCGGATC-----GAGCGCGC 282514
Qy 315 pAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTy 335
Db 282513 CGAC-----CACACACGCGC-----ATCGAGGTCCACAGACCAAGAT 282475
Qy 335 rGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLe 355
Db 282474 CAACTGCAAGGGGAGTACCGTCCCGATGCGCGCGGATCGACAGATCCGGTTCGGGC 282415
Qy 355 uAsnArgAsnAspTyrHisHisAspGlu-----LysAlaGl 367
Db 282414 CGCGCGGACCGACTACCGGCACACAGATCGGCTTCGCGATCCCGCGGACCCCAACAG 282355
Qy 367 yAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHi 387
Db 282354 CGACGCGGTGCGCGAGACCTTTTACCACCAAGAGAGAGAGATCCCGACCGGTGCACT 282295
Qy 387 sGlnPro-----IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGl 403

```

```

Db 282294 GATCGCTTCAACGCGCGCTTCGCGAGTTCACGAGCGGTGGGCTTCCAGTCCGCGCA 282235
Qy 403 yGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAs 423
Db 282234 CCAGAACTGACCGCGCGAGCCCGGACAAATCCCGGACGCTGTTCAACGCGCTTGGGA 282175
Qy 423 p-----AsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTr 439
Db 282174 CCCCACACACACGCGGTTCGCTTC----- 282138
Qy 439 pAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTy 459
Db 282137 -AACGAGTTCAAGTTCAACGCGACCGCGCGCGAG---ATCGCGCGCGCATCGACGA 282082
Qy 459 rAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAl 479
Db 282081 TGTGAGCTGACGCGCACGACCGCCGAATTCGCGCGGACCTTCCTGCGCGGAC----- 282030
Qy 479 aHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHi 499
Db 282029 -----GGACGCGCGAGTTCGCGCAATCGCTCTACACGCGCAAGAG 281980
Qy 499 sLysLeu-----SerLeuThrAlaSe 506
Db 281979 CGGACGATCGGCTGTTCAGGACCTGCGCGCGGATGTCGCGACGATCACCGCGCA 281920
Qy 506 rHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaTh 526
Db 281919 ATATGTCGAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 281860
Qy 526 rAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLe 546
Db 281859 CGCGAGTTCGACATCGGCAATCCGAACTGACGATCGAGACCGCAAGTTCGTGAGCT 281800
Qy 546 uAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPh 566
Db 281799 CGCGGTGCGCAAGGCGCGGCGCGGCTCGGTCGAGGCGGCGGCTACTACACGCAATT 281740
Qy 566 eGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAs 586
Db 281739 CGCAATTTTCATCTTCGCGCGCTCACCGCGGTGATGTCGCGC-----GA 281695
Qy 586 pAsp-----SerGluMetLysLeuValArgTyrAsnGlnSe 598
Db 281694 CGACTTCGCTCGTGGCGCACCGCGCGCGCGGCGGCGGCGGCTACTCCGCGCG 281635
Qy 598 rGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrAr 618
Db 281634 GAACGGCAATTCGTTGGCGGCGGAAATTCAGTCGCGAGCTC---GACGTGCGCGCGCTCCA 281578
Qy 618 gIleGlyValSerGly-----AspTyrValArgGlyArgLeuLysAsnLe 633
Db 281577 GCGCGCATCTGGGGATCGAGAACCGAGCTCGAGTGTTCGCGGCGGCGCTTC----- 281526
Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGluAspAs 653
Db 281525 -----ACCGCGG 281518
Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 281517 CACCAAGTTCGCGGATTCGCGGCTCAGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 281470
Qy 673 rAspArgIleAsp-----AlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLys 690
Db 281469 CTGCGCGGACGACCAACTGGCTGATGCGCTCAACTGTTGCGCGCTTCGCGCGAGACAA 281410
Qy 690 sLeuAlaArgTyr---GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAs 709
Db 281409 CGTGGCGGTGATCGCGAGACCGCGGAGCTACACCTCTCTGAGCGCGGAGGTGAG 281350
Qy 709 nTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyr----- 722

```



```

/translation="WERSRSIFFIMLRKTNFQGLNLSYIDTNSKSKPTILLCHANCY
SAFTYKVESLONSRYVALDPAGHGESDSTLNFDRWYFRDQVLSLSENNVNI
GHSGLGASLLASYSRSPDKKVIADHPDPTLNFLOITYSRIFHFNPLAQVAIKRRE
FKNDLVKKIYKRTSPRMDGDIYDDIYQSCFRLGVNGEAILRCAFEVBAKIPDSVS
YRSIFQKIKITETHITIPDPHEVCSPNTAKKIISGNIRSSLEIWPKTSHFFPFEEKH
KTLDRILFVLD"
6714..7046
/locus_tag="LA3148"
6714..7046
/locus_tag="LA3148"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:24197045"
/translation="WILQIRKSGFLSILLIIFVSSGLHIIHVEQEIQKFSGYSTPNKD
TKQIVACICQFQRESNLNPNFLTGNQPSVFLKKIRISFVIFIHFNPNQRFGRBA
PPFLUPLS"
7092..9425
/locus_tag="LA3149"
7092..9425
/locus_tag="LA3149"
/codon_start=1
/transl_table=11
/product="hemin receptor"
/db_xref="GI:24197046"
/translation="MKKYLIFLTLILFCQNEKNNQDLMLRLLANTFNETRSDCV
YCTDQAFQNGCSFSPNIPWSCQGYSSGRQKNSIRISCEDLTSKGWFPENSQQA
KSCPSLSCPPEAYRAAFTPDGF"
complement(9855..10016)
/locus_tag="LA3151"
complement(9855..10016)
/locus_tag="LA3151"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:24197048"
/translation="WNSYFKIFPSRFLTNSRYLELLKNFIAVNKNKIASIYFFHTEI
DVEFTFOKL"
BASE COUNT 3177 a 1970 c 1645 g 3422 t
ORIGIN

```

Alignment Scores:

Pred. No.:	3,286-14	Length:	10214
Score:	396.00	Matches:	174
Percent Similarity:	39.95%	Conservative:	130
Best Local Similarity:	22.86%	Mismatches:	303
Query Match:	9.81%	Indels:	155

```

DB: 1 Gaps: 29
US-09-936-377-2 (1-758) x AB011477 (1-10214)
Qy 28 GluGlnSerValGlyLeuGluThrValThrValValValGlySerArgProArgAlaThr 47
Db 7404 GAAAGTTCTTGGACCAATCCGCGATCAACGTCACCGCTAAATCC-----ACA 7451
Qy 48 SerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeu--- 66
Db 7452 AATTTCAGATTTTAACTGCTCCACACCAATACCGTTTTCGGTAGACAAATAGAC 7511
Qy 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
Db 7512 CGTCAAGGGGAGAAATCGCAATGTCTATCAACAATCTCCGGAGTCTTCCACACCT 7571
Qy 87 GlnTyGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgGlyLeuLys 106
Db 7572 ACAACAGGATCAGAACTTCTAAACCGATCATCCGAGGCTTACCAGCAAAAGGGTTTA 7631
Qy 107 ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
Db 7632 GTTATGACC---GACGGTATTCTGTCAGAGAACAACAATTTGGTCACGACCACAGTA 7688
Qy 127 MetValAspThrAlaLeuSerGlnGlnValGluLeuArgGlyProValThrLeuLeu 146
Db 7689 GAATTAGATCTTTCAATATTCAAAAATTTGAAATTTAAGAGGCCCGGTAGTTTACTC 7748
Qy 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro----- 164
Db 7749 TAGCGCTCAGACCTCTTGGAGGTGTAGTAATGTCTATCGAGACAACACCTCTTCC 7808
Qy 165 ---GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLysSerSerGly 183
Db 7809 GGAGAAGGAATCCAAAG-----ATGCGAGGCAATTTTCAATTCCAACAGTATTCTAAC 7862
Qy 184 AsnLeuGluLysLeuThrSerGlyGlyIleAsnIle---GlyLeuGlyLysAsnPheVal 202
Db 7863 AAC-----AAACAGATGCTGCACTTGTCAATCTACGAAA--TCTAAACGGTTTCG 7915
Qy 203 LeuHisThrGluGlyLeuTyArgLysSerGly---AspTyrAlaValProArg----- 219
Db 7916 TTATCGGTAGCTCCCAATTCAGAAAGCGGGTAAATTTACTACTCCGAACGGAACCT 7975
Qy 220 -----TyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGln 237
Db 7976 TCGTAATACAGGATGATAGAAAAAATCAAGTGGTCCATTGGTTTGACCGGGGATG 8035
Qy 238 HisArgAlaValLeuGlyTrp-----ArgLysArgPheTyArgArgThr 252
Db 8036 GGGTAATTTTATTTGATTTCTTTTAAAGGGAACAACCAAGATCTTTTACACAA--- 8092
Qy 253 TyrSerAspArgAsp---GlnTyGlyLeuProAlaHisSer-HisGluTyArgAsp 271
Db 8093 ---TCCAAACGAAATCTCTGGAGCTACTGTCTCTCAAAACATTTCTTACGAGAAATCTCA 8149
Qy 271 pCysHisAla-----AspIleIleTrpGlnTy 280
Db 8150 TTTTCATCTGTTTTTATTTTTCGCGAGAAACTTGGAACTTGATTTTCTCTATCAAG 8209
Qy 280 sSerLeuIleAsnLysArgTyLeuGlnLeuTyProHisLeuLeuThrGluGluAspVa 300
Db 8210 AAAC-----AATCGAAGAGAAATCGAATCTAAAAATAAATCTCTCAATTCAGATAT 8263
Qy 300 lAspTyArgAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHi 320
Db 8264 TCTTTTG-----GACGAAATCGAGAAAGTT--- 8288
Qy 320 sAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyGluLeuArgAlaGl 340
Db 8289 -----TTAGATAAATCTTTTTCAGTTCTATCAAGTACTTCCAA 8326
Qy 340 uTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360

```

```

Db      8327  AGAAAAA-----TACCAAGGTTTAAACCTTTTCTTAGAC----- 8360
Qy      360  rHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnGlnThrGlnAs 380
Db      8361  -----ACCACATAC 8368
Qy      380  nAlaAArgLleGluLeuAArgHisGlnProIle---GlyArgLeuLysGlySerTrpGlyVa 399
Db      8369  AGCGGAGCGTAAATTCACCAATAAACCCTATATCAATTTTGAAGAAGAACTATCGGAAT 8428
Qy      399  lGlnTrpLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnPr 419
Db      8429  ATCTGCCTCGAACAAGAAATAAACAATCGGTACA-----GAGCC 8470
Qy      419  oMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTr 439
Db      8471  GTTGATTCCTCTACGCTATATTAAACATTCGCGGTATTTTGAAGAAGAACTTAAACT 8530
Qy      439  pAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTy 459
Db      8531  GGGTTCCTTAACCTTTCAGCGCTGGAATTAGAGGTGATAAACGTTTCGTCGACATTCGAAA 8590
Qy      459  rAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAl 479
Db      8591  TAACCAACC-----CTTGCTGT 8608
Qy      479  aHisAArgGlnThrAlaAArgSerPheAlaLeuSerGlyAsn-----TrpTyrPh 495
Db      8609  CTTTCGAACAACCAAAATTTATATATGACACCGGTTCCACCGGTGTGTAGGAGAT 8668
Qy      495  eThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGl 515
Db      8669  CTATAATCTTTTCCACCATTTTAAATACGACGCGGTTTAGGCGTCTACTCCTTT 8728
Qy      515  nGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHi 535
Db      8729  TGAGCTTTTCTCAGCGAGTCCACGAAGAACCGGTAAATTTGAAATTTGGAACAATCA 8788
Qy      535  sLeuAsnLysGluAArgSerAsnLysLeuGluLeuAlaLeuGlyTyrGluGlyAspArgTr 555
Db      8789  TCTAAAACCTGAATTAATCAATATCTGGATTTTCTCTAGATTTGCTTCTTAAAT 8848
Qy      555  pGlnTyrAsnLeuAlaLeuTyrArgAsnAArgPheGlyAsnTyrIleTyrAlaGlnThrLe 575
Db      8849  TCAAACTGAGATAGCTGTTTCAAAATCATATTCAAATTTCAATATGCTCAAGTAT 8908
Qy      575  uAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTy 595
Db      8909  C-----GCCCAATAGACTTGGATTTCTGTTTACCAAAATACGAGTA 8950
Qy      595  rAsnGlnSerGlyAlaAspPheTyrGlyAlaGlu-----GlyGluIleTyrPh 611
Db      8951  TAAACAGGAACCGGATTTTAAAGAGGAGATTTTCGATTCAGACGAACTTTCCG 9010
Qy      611  eLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuTy 631
Db      9011  TAAACTCGTT-----TTTTCGCGTGAATCGATTTGTTTATCTAGAAATCAA----- 9059
Qy      631  sAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnAArgProPheIleAlaGl 651
Db      9060  -----AG 9061
Qy      651  nAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSe 671
Db      9062  CGATCTCATCTCTCTAGAACTACACCAAAATCGACGCGCTGCTTAGATGGAC 9121
Qy      671  rLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal-----Ph 686
Db      9122  GGAAGATTCTATTTAGGTTTAAAAACTTCTACTTTTCCATCAACGGTAGGTTTATGA 9181
Qy      686  eAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLe 706

```

```

Db      9182  CTCTCAATATAGAGTCGATCTCTAAAGAAACACCTACCAAGGTTATATCTTATAGACGT 9241
Qy      706  uGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTyrAsnTyrTyrValLysAlaAs 726
Db      9242  TGGACTCGGTTTGAACCTT---CCTCATTTCCGAGACGGAACCTTCAAGGCTACTGTAGA 9298
Qy      726  p-----AsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAs 742
Db      9299  CTTGAGTATTCAAAATGATTAATCAATGTTCTTATGTAGATCAATCTCAGTAGATATAAGGA 9358
Qy      742  p 742
Db      9359  C 9359

RESULT 27
LOCUS   AE016931/c
DEFINITION Bacteroides thetaiotaomicron VPI-5482, section 6 of 21 of the complete genome.
ACCESSION AE016931 AE015928
VERSION   AE016931.1 GI:29338516
KEYWORDS
SOURCE
ORGANISM Bacteroides thetaiotaomicron VPI-5482
          Bacteroides thetaiotaomicron VPI-5482
          Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
          Bacteroidaceae; Bacteroides.
REFERENCE 1 (bases 1 to 303750)
          Xu,J., Bjursell,M.K., Himrod,J., Deng,S., Carmichael,L.K.,
          Chiang,H.C., Hooper,L.V. and Gordon,J.I.
          A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis
          Science 299 (5615), 2074-2076 (2003)
          12663928
          PUBMED
REFERENCE 2 (bases 1 to 303750)
          Xu,J., Bjursell,M.K., Himrod,J., Deng,S., Carmichael,L.K.,
          Chiang,H.C., Hooper,L.V. and Gordon,J.I.
          Direct Submission
          Submitted (31-OCT-2002) Department of Molecular Biology and
          Pharmacology, Washington University in St. Louis, 660 S. Euclid,
          St. Louis, MO 63110, USA
FEATURES             location/Qualifiers
     source
     1..303750
        /organism="Bacteroides thetaiotaomicron VPI-5482"
        /mol_type="genomic DNA"
        /strain="VPI-5482"
        /db_xref="taxon:226186"
        complement(87..1649)
        /locus_tag="BT1210"
        complement(87..1649)
        /locus_tag="BT1210"
        /codon_start=1
        /transl_table=11
        /product="cytochrome D ubiquinol oxidase subunit I"
        /protein_id="AA076317.1"
        /db_xref="GI:29338517"
        /translation="MIESIDSLIDWSRAQPMATMHWIFVPLTGLAVVMGMETIL
        YKGNFQFKETAKFWKLGINFAIGVATGLILEFEGTNWSYSPFVGDFGAPLA
        IEGIALFMEATFIATVFGWQKVSQFHLASTWLTGLGATISAWWILVANMMQHPV
        GMEFNPTRVEMVDFAVATPVAVNKFHTVLSGWLGVGVVGVSCWFLKKRRR
        EFALASIKIGALFGLVSLAVTGDGSGVQIACTQPMKLAAVVGLVEGGTVNGLVGI
        GMLNDEKTYNDGKDPFLRIEIPSMLSFLAERVDGVVPGITWNIIEGGYQKDGTTA
        LSAEKIERGATGALAAAYAAKASAGHEEDAAYQVLQENMEFYGGYTKQVNLV
        PNFNFYAFRNVILGGYFLFFIVLFFYKKDKLSKRMWMMHIALTLTPLAYIAQG
        AGWVAVCGCGPWAIRDMLPTSAAISKLDVGSVQATFFILFLFTVMLIAEIGIMVKA
        IKKGPEEISNR"
        complement(1670..1906)
        /locus_tag="BT1211"
        complement(1670..1906)
        /locus_tag="BT1211"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AA076318.1"
        /db_xref="GI:29338518"

```


Score:	392.00	Matches:	190
Percent Similarity:	37.82%	Conservative:	122
Best Local Similarity:	23.03%	Mismatches:	311
Query Match:	9.71%	Indels:	202
DB:	1	Gaps:	36
US-09-936-377-2 (1-758) x AE016931 (1-303750)			
Qy	14	LeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeu	33
Db	248987	TTGTTTGAACCTTACCGCTCCGACGACAGAGTCCGATCAATCAAGAGATGTTTTT	248928
Qy	34	GluThrValThrValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThr	53
Db	248927	CCTGATGTTGTTGACCGAGTCGTACCAACAGCGCTCAAGCCAAG	248880
Qy	54	SerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAlaVal	72
Db	248879	TCTGCATTTACGGTTGAAGTGATAGAACAGAGATTTCTCGCAACATTTTACGGGAAC	248820
Qy	73	LeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnThrGlyGlyAla	92
Db	248819	TTCATCCAGCAGTGAGAACATTCGCGAGTACAGCTATGACATAGGCTCCGGCTTC	248760
Qy	93	SerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisGly	112
Db	248759	TCAAAACCGATGATTCGGGGGATGGGATTCACACCGGATAGCGGTATTG	248703
Qy	113	GluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu	132
Db	248702	ATCAACGAGAGCCGACGAATGGGAGCGCCAGCGACTGGAACCTGATCCCTTCAT	248643
Qy	133	SerGlnGlnValIleuArgGlyProValThrLeuLeuLysSerSerGlyAsnVal	152
Db	248642	ATCGAAGCAGTCAATGCTTGAAGTCTCTCTCACTGCTTTACGGCAGTATGCCATG	248583
Qy	153	AlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGlyVal	172
Db	248582	GGTGGAGTGAATGATAGCTCTCTCCATACCTCT	248532
Qy	173	SerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGly	192
Db	248531	TTCCGGCGATCTACCCCTATTGGGCAATCCGTCAAC	248478
Qy	193	IleAsnIleGlyLeuGlyLysAsnPhe	208
Db	248477	CTGATCTGGGACTTAAAGAAATGCTGTATGCGCAAAATCCGTTATTCCGAA	248424
Qy	209	TyrArgLysSerGlyAspTyrAlaValProArg	224
Db	248423	---CAGCATTCTCGTGATTACCGCATCCCTACGGATACCATTTGCTACCTGACT	248370
Qy	225	ArgLeuProAspSerProArgArgPheAlaAsn	241
Db	248369	AAAATGCCCTCATGCGGCTAAATGAAATATCGCGAGGAGTGGAAACCAATATCCGC	248310
Qy	242	LeuGlyTyrArgLysArgPheTyrArgThrTyrSerAspArg	257
Db	248309	CTTTTACCAATATCAGCGAAAGGCTATAGACAGACTATTCATCATGATGCTAT	248250
Qy	258	AspGlnTyrGlyLeuProAlaHisSerHisGluTyrAsp	270
Db	248249	CAGAAAGCGGGTTCTTTCCGGGCGCACACGGTGTCCCGACGATCCGCGGTAGAGAT	248190
Qy	271	Asp	271
Db	248189	GACGAGACAGCCGGAATATCGAATCTTACGACGAGTAATCATCTGAAAGTAACC	248130
Qy	272	CysHisAlaAspIleIletrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyr	291
Db	248129	ACACATCAGCAATATGCTTGGGAGAA	248103
Qy	292	ProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPhe	311
Db	248102	---CTGATCTCTCGGTGATTGGGATTTTCAGATAAC	248067
Qy	312	HisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProThrIleAspLeuArg	331
Db	248066	CATCGGAGGAATGGAGTCTTTTCATCTACTACCTACCGC	248028
Qy	332	AsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeu	351
Db	248027	---TCCAGCCTCGCGC	248013
Qy	352	ArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGlu	371
Db	248012	---GAGAAGACCCGCGCAAA	247992
Qy	372	AsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGly	391
Db	247991	CTGGCTTTTCATCTGAACATACATGCTTCGTAAAGCCCGC	247947
Qy	392	ArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAla	411
Db	247946	---TTTATTGGC	247920
Qy	412	ThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhe	431
Db	247919	ACGTTGGGGTGGATAGCAACATCAG	247866
Qy	432	PheGlyValGluGlnAlaAsn	439
Db	247865	TTGTTGGCGGAGTACCATTCTCTAGTACCGGATTCCTGGCTTACGACGTATAAGCCT	247806
Qy	440	AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIle	457
Db	247805	AATAACATCTCTCTGTCAGTGGGCGTGGATATGATGTTATATATGTCATCTCT	247746
Qy	458	ArgTyrAspLysAlaLeuLeuLeuAsp	473
Db	247745	CCGACACGAGGATGTTATCTGCGGATACCTTCGGAGGCAAGATATGATGACAA	247686
Qy	474	LeuProAspLeuGlyAlaHisArgGlnThrAlaArg	487
Db	247685	GTAGAGTCTTATAATGAACAGTCAATCGCTCCGAAAGCATTTTCGGTGATATTCCTTT	247626
Qy	488	AlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis	507
Db	247625	TCTTTGGACGTGGATGGACACCTCCGTCAGCACATGTTGAAGTGAACATCGGACGT	247566
Qy	508	GlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn	527
Db	247565	AGCTTCCGCTCCCGGAGCAATGAATGCTGCTAAACGCTGATCATCAGGCACATTC	247506
Qy	528	ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAla	547
Db	247505	CGTCACGACGAGGAGATGCTTCTGAAATCCGACGAGATGCGAGATGCTGCTCC	247446
Qy	548	LeuGlyTyrGluGlyAspArgTyrPheGlnTyrAsnLeuAlaLeuTyrArgAsnArg	565
Db	247445	TACAACCTCGCATATCAGCA	247392
Qy	566	PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGlu	585
Db	247391	TTCACTACTACATCTTTTCGCTCC	247344
Qy	586	AspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla	605
Db	247343	CCGATGCGCA	247287
Qy	606	GluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyr	625
Db	247286	GAGCTACGATAGATTAATCTCTGCTGCTCACTTACCTGCTTCTCCGAGATAAC	247227
Qy	626	ValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsn	645

Db	247226	GTC-----	247224	gene
QY	646	ArgProPheIleAlaGlnAspGlnAsnAlaProArgValProAlaAlaArgLeuGly	665	CDS
Db	247223	-----TACACCTACAACTGTGACGAGCATATTCG-----TTAAGT	247188	
QY	666	PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal	685	
Db	247187	TTTTCCTCCGCTCGGTATGCGGAATACGTTGACATGGCAAGAAACACTATATGCTT	247128	
QY	686	PheAla-----GlnAsnLysLeuAlaArgTyrGluThrArgThr	698	
Db	247127	TATGCGGAGTGGCAACTATTGCCCCGACGACCGGTAGACCGTATGAGGACCGTACA	247068	
QY	699	ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGlu	718	
Db	247067	CCGGGGCGCTAACCTGTTTCATTTGGGTGGCTCGCTG-----AATATCCGGTTGGCAGG	247014	
QY	719	TrpAsnTrpTyr-----ValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAla	735	
Db	247013	ACTAACGAGATGAATAACCTGACCGCCGCTAATATCTTCAATACCGATATTACAAT	246954	
QY	736	HisSerSerPheLeuSer-----AspThrProGlnMetGlyArgSerPheThrGlyGly	753	
Db	246953	CATCTTAGTTTITACAGGAAGTAGAAATCCGGACCGGACCAACTCCAAATTATTA	246894	
QY	754	ValAsnValLysPhe 758		
Db	246893	ATCAAAATTCATTT 246879		
RESULT 28				
AE006167				
LOCUS	AE006167	11549 bp DNA linear BCT 08-MAR-2001		
DEFINITION	Pasteurella multocida PM70 section 134 of 204 of the complete genome.			
ACCESSION	AE006167	AE004439		
VERSION	AE006167.1	GI:12721638		
KEYWORDS	Pasteurella multocida			
SOURCE	Pasteurella multocida			
ORGANISM	Pasteurella multocida			
REFERENCE	1 (bases 1 to 11549)			
AUTHORS	May, B. J., Zhang, Q., Li, L. L., Paustian, M. L., Whittam, T. S. and Kapur, V.			
TITLE	Complete genomic sequence of Pasteurella multocida, Pm70			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)			
MEDLINE	21145866			
PUBMED	11248100			
REFERENCE	2 (bases 1 to 11549)			
AUTHORS	Zhang, Q. and Kapur, V.			
TITLE	Direct Submission			
JOURNAL	Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA			
FEATURES	Location/Qualifiers			
source	1. 11549			
	/organism="Pasteurella multocida"			
	/mol_type="genomic DNA"			
	/strain="PM70"			
	/db_xref="taxon:747"			
	77. 391			
	/gene="PM1279"			
	77. 391			
	/gene="PM1279"			
	/codon_start=1			
	/transl_table=11			
	/product="unknown"			
	/protein_id="AAK03363.1"			
	/db_xref="GI:12721639"			
	/translation="MFAMLVEINIKEGKEQBFLEVFARNHIGTROEPGNLRFVLRDP			
	RIKTRFYAYEVINEQALEAHKQTAHYHRCVAELEPLMTGARSKKLFDWVMPENLTNE			
	CG"			
gene				
CDS				
384. .1073				
/gene="dod_1"				
/note="synonym: PM1280"				
384. .1073				
/gene="dod_1"				
/codon_start=1				
/transl_table=11				
/product="Dod"				
/protein_id="AAK03364.1"				
/db_xref="GI:12721640"				
/translation="MDKQIIOQLTKPTQPIISVGILASDWLKFADTLTTLSRHHRLRLHP				
DIGDQFSFFFTGALAVKQFSPMLKQVHLMVNDPFFHAKACADAGADITIQEQBT				
DCIAETISVREHYPLDILGLCPTEIDLLTPYLAQVDLQIOLTLDRPTGVKAEVD				
AVIKRITRISNMLGEHRDQKLISVDGSMNLALASQLFPLGIDWVSSALFSQADVDA				
TLSEKTHLCR"				
complement (1122. .1997)				
/gene="PM1281"				
complement (1122. .1997)				
/gene="PM1281"				
/codon_start=1				
/transl_table=11				
/product="unknown"				
/protein_id="AAK03365.1"				
/db_xref="GI:12721641"				
/translation="MIKGILLSLISASLVFGYLYFYFTLQPLGGEDIFGFRIIFTPLF				
VIAAVLPFKQAVGQYFVRIKQKQPLIFGFLFNSAMMGIOIWLFLWPNNGSALSVS				
FGYLLPLTMVLVGLVRFKEHTSRVKFLAVTAAIGVFNNIILKGLSWEALVSRGY				
STVFATRKILKINDLAGCLESLLPLVCIFYPAWQVDLQAVQANENILLLLVLLGLI				
SGVALNTYIVASSLLPINVLGUGTAEPIIMLVFSLFGLQGLDSEITIPFLICLMISMI				
LFWSEGLVRLKRRYK"				
2144. .4480				
/gene="PM1282"				
2144. .4480				
/gene="PM1282"				
/codon_start=1				
/transl_table=11				
/product="unknown"				
/protein_id="AAK03366.1"				
/db_xref="GI:12721642"				
/translation="MLCKPNHGLLILISEVVFADTQDLDEIHVHVSVNKGKSLTALA				
SOVRISKLDNVKSGTIGETITNVSGVNNYFGPNAGRPIMIRGLSGQVQIMQNDMP				
IHDMAASNSNLAISANLASVKEIQIKSGVASVLYGKSGIGGAVNLTDIAIQQHTAEK				
AFRGVEWRAGLNAFTYQVQLQNNQNVFAYILDASLSEISRRKIPGRSKDQICVDY				
EYLKNTWKLRSQCQNIKLSINPKFYIKIQFYLDHYKNPDYGLSEGDOYTNMSAG				
GKPNLPEYVGSPPYAEKPGLEVPVPGNGRIINAHSSTKOLTGVASVLDKGYTG				
ASFHENTRYGVGSPSYLASRAASQKSVLPVSVNTEINLNKQLNALVMPISKVE				
IOLGQKSRDEKLEWLSHLSNAFTYNDYRSSFEHQPLFNHHVLTGIGVEMNRQKILFS				
SGKDSYLVNMDTKSLFILEKIDVSPFVVEFGQWQVKVYVVDITNHISSRGKGGY				
SKNTFFVONTHLGLLFOVDTWYLLKQRSQWERAPEVNELYADNNHYALLIEENGDG				
RLKKEQSDNWLSTGVNLEQFNAELTFQRFNDFNPHLYGTGTSRAGLAVKEWROAPL				
RMNGVELDLTYOHTPHFGDLTWHLFADYVHHKLTAPDHIOTKRIQYPLHPTSRIG				
GDLTVMQWRGFTSVTHVVKOKRVTKVDEVDDELPAKKETLVDLGLSYVYPLTRATLEI				
SINIONLTNRTEARLNNSPRLAPLPGRNVLGVKLMF"				
complement (4540. .5421)				
/gene="ilvy"				
/note="synonym: PM1283"				
complement (4540. .5421)				
/gene="ilvy"				
/codon_start=1				
/transl_table=11				
/product="ilvy"				
/protein_id="AAK03367.1"				
/db_xref="GI:12721643"				
/translation="MNFYTLKFLHLCETKNFAKTAONHMSPTLSRQIORMEODVG				
QILFLRNNRQVELTDAGEKLLPEAQCSWMOQLKQQLRDSQAQELNGELKFLCSVTA				
AYSHLPAILERTRTPYKVEIQLTTGDSLSALQVETQVDLAGPKNLPHNIAFH				
YIDISLIAPIACSATQLLQQTPIINWQIPFLPVEGPARQIDQWFOQKTKHP				
KIYATVAGHEAIMPWALCGVAMPDPVVIKNMSQVSIHLHKTPTTTPFALGICVQ				
KKSLTQPLVRAFWQLQ"				
5674. .7149				
/gene="ilvc"				
/note="synonym: PM1284"				
5674. .7149				

Db	4028	GTAGATTGATTAACTTATCATCAACATCAAACTCCCATTTGGTGATTTGCATGCAT	4087
Qy	621	ValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer	639
Db	4088	CTTTTGGCTGATTATGTTTCATCAAAATTTGACATTAGCGCCATCATATTGATGACGAA	4147
Qy	640	GlusAlaTyrGlyAsnArgProPheIleAlaGlnAspGlnAsnAlaProArgVal	659
Db	4148	AGAATTGGTTAT	4171
Qy	660	ProAlaAlaArgLeuGlyPheHisLeuLys	670
Db	4172	CCAACGCTCTAGGATTGGGGAGATTTCACAGTACAGTGGCAACGTTGGAGAGATTACT	4231
Qy	671	SerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLys	690
Db	4232	TCTGTGACA	4270
Qy	691	LeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyr	710
Db	4271	GAAGTCGATGATGAATTCACGACAAAAAATTTACGTTAGTGGATTAGTTTACTTAT	4330
Qy	711	ArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsn	730
Db	4331	GTTTATCCGCTTACTCGACGACACATTAGAGATATCTTTGAATATCCAAAAATCTCACGAAT	4390
Qy	731	GlnSerValTyrAlaHisSerSer	747
Db	4391	AGGAAGCAAGATTGAATAACTCACCTTACGTTCTCTCGCGCTTTACCT	4444
Qy	748	ArgSerPheThrGlyGlyValAsnValLysPhe	758
Db	4445	AGAAATGCTACGTTTAGTGTAATTCATGATGTTT	4477
RESULT 29	AE004349	17725 bp DNA linear	BCT 31-JUN-2000
LOCUS	Vibrio cholerae chromosome II, section 6 of 93 of the complete		
DEFINITION	chromosome.		
ACCESSION	AE004349		
VERSION	AE004349.1		
KEYWORDS	GI:9657434		
SOURCE	Vibrio cholerae		
ORGANISM	Vibrio cholerae		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;		
AUTHORS	1. (bases 1 to 17725)		
	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,		
	Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,		
	Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,		
	Emmolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,		
	Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D.,		
	Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,		
	Mekalanos, J.J., Venter, J.C. and Fraser, C.M.		
	Direct Submission		
TITLE	DNA sequence of both chromosomes of the cholera pathogen Vibrio		
JOURNAL	choleræ		
MEDLINE	Nature 406 (6795), 477-483 (2000)		
PUBMED	20406833		
REFERENCE	10952301		
AUTHORS	2. (bases 1 to 17725)		
	Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,		
	Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,		
	Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,		
	Emmolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,		
	Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D.,		
	Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,		
	Mekalanos, J.J., Venter, J.C. and Fraser, C.M.		
	Direct Submission		
TITLE	Submitted (14-JUN-2000) The Institute for Genomic Research, 9712		
JOURNAL	Medical Center Dr, Rockville, MD 20850, USA		
FEATURES	Location/Qualifiers		
source	1. 17725		
	/organism="Vibrio cholerae"		

```

/mol_type="genomic DNA"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="II"
/notes="biotype: El Tor"
/complement (90..1652)
/gene="VCA0052"
/complement (90..1652)
/gene="VCA0052"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95966.1"
/db_xref="GI:9657435"
/translation="MRLKPSSTNLSCYVQATQDGLVEPEGEIIVNDQLPPDARGL
QLNCKYTLACNFCGLSEKHYVLQVNPKPAMVCRBCGAPPLANNRQVNVNHLRL
QLHSDGILPACHNPACPTGLSVTHKQIYHAFVSGDRQRYRCKTCHSTFVQWGAN
HKLNFBSLLGLFTGYSVREICRKLISNPNTFYDHLDHIAASCRRLKLMIDARWNH
AKAYQFASHYQLOAHSHNGVYWTADQATGYILCOHLYNSADEQPTGSLDHPYQT
PARFVQYAAEAQATPHKALTLLRERIDNRQYLILARANVEDPMGNLAQHYPSKGA
VIRPPYTSYAHVHLVDMCDPGKRSIFMPQDPLLRSAALSLSLRIRCONVLMYVE
BDPMQASPLEKTDIVHMGWRDRWIAHSGTKAGICVLAGSNPDPOHLOWASIR
QLEYQNEFQLLESFINEPRKLRPGCIQLPLDIFRAMEHLCYQDKHGTAQNLGL
SQRLTLKELLS"
/complement (1624..2361)
/gene="VCA0053"
/complement (1624..2361)
/gene="VCA0053"
/notes="similar to GB:U14003 SP:P09743 GB:M60917 PID:147309
PID:537224; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="purine nucleoside phosphorylase"
/protein_id="AAF95967.1"
/db_xref="GI:9657436"
/translation="MAKLNWRVTWATPHINAQPDGFAETVLMPCDPLRAKYIAETPLF
DVKQVDVMSVFGFTGYTKGKVSVMGHMGIPSCSYVHELIAEYGVKNIIRIGSGG
AVRDDVKLMDVIGMGASTDSKNVRIRPSGHDFAIADYDLLETAVNQARAQVPPKV
GNVFDADLYTPETPEIFEMKKIGILGVDMEAGIYVGAADLGARALTILTVDHILR
GEKLSSEDRQKSFNDMMKVALETAINI"
/complement (2420..2599)
/gene="VCA0054"
/complement (2420..2599)
/gene="VCA0054"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95968.1"
/db_xref="GI:9657437"
/translation="MILYCTNIFVQNFPTPTKRLAQRLPYQFQISSLSTAYPDFKR
EYFVTLSSSLVDIG"
/2641..3609
/gene="VCA0055"
/2641..3609
/gene="VCA0055"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95969.1"
/db_xref="GI:9657438"
/translation="MLETMRQITPLKIGISACLIGAKVRFQDGHKISHFVTNELDRYA
EFVSPCMGMLPVPRPTRLISENERIALVETKDSRSDHAALESYSQOKVAELOH
AELCGYIVCAKSPGCMERVKYHNGTEKDGVGVTYTRILMEKMWLPFIEDGLNDP
VLRENYLFEVCLHDFYTSMDGAPSAKKIVFEHSEYKLTMAHNPQYALGRILVAVY
KQYLDSTFQETRIGLMALAHASRKNNTVMHLQGFYKRSKQKQAEIAQVIES
YVGBELPLIAPITLLKHLYALFPDADYLAQQRYLNPYPOELKRLRYGL"
/3599..4396
/gene="VCA0056"
/3599..4396

```

```

/gene="VCA0056"
/notes="similar to SP:P33358 PID:405858 GB:U00096
PID:1789448; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, MerR family"
/protein_id="AAF95970.1"
/db_xref="GI:9657439"
/translation="MVCDEKRYAIRVAEITGVKPVTLRAWORRYNLVOPDTEKGHR
LFTDQDIEMIRQISWLAGVAIGKVGALLQSGVSESIIPQVQLECEETLLTALA
ALORSKAQIATVLKYEPLSMQAFQVPVTEALERVKGLRSHQIGLFRLLMSKL
AFILDAENKAIVKGCICISLDEAGSLNMLWALAWENGHOVTLLEAVDDIRGLEEN
PGLAQYILALHAHRALPAACQSALASLQQQGEQCIVLSNVLLQQLQS"
/4406..5815
/gene="VCA0057"
/4406..5815
/gene="VCA0057"
/notes="similar to GB:K01299 SP:P00914 PID:147268 GB:U00096
PID:1651311; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="deoxyribodipyrimidine photolyase"
/protein_id="AAF95971.1"
/db_xref="GI:9657440"
/translation="MRLVFERDLRSFNTALTAALNSGDPVAAVYIATPEOWHOHL
APIQADLWERLAELOQLAALNVPLFYQVADFOAAVAVSOLAKTINATOVLANRD
YELDEQQRDQLAQQLISQGIWISAFDDKCVLPQGSVRTKQGEPEFKVTPFKRAMLTL
FQPPVIGKRPVALWNVPSALAEVLVHPEQAFDPRIDSTPWAADFETVRAQDRFCR
ERVQDYHQARDPFAREGTSSLSPLAIGVLSARQCVRALYHSESGEAGQVWLSE
LWRFYFQHLVAIEPNLSKSRDFVEWGRLEWNNDEKFLQMCBKGTGYPYVDAAMRQ
LNOTGMNLRMIVASFLTLDLHDMWGERYFMSRLIDGIDYAAANGGQWQCASTGC
DQQYFRIFNPVSQGEKPDGPDGDFIRRWVPELRSVSSAVIHQPMTPYPAVNSVLYPARL
VHKQREVTLLRYKTAKG"
/5824..6744
/gene="VCA0058"
/5824..6744
/gene="VCA0058"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95972.1"
/db_xref="GI:9657441"
/translation="MMWKVWITGALAVSWSINAAVFDLPBEGSSIVGNTQYHKVEKG
ETLADIATKYDIFGLMAANBGVDFFLPQEGFVISIPAKIILPKVAVGIVINLAEL
RLYFEPNEKGVHLPVGGIRGRDTPVMQTSISSKRYPTWTPPASIRKEYKAGID
LPPVVPAGPENLGDYAMRLAYSGSEYLLHGNKDFGCGMRVSAGCIBMDPKOLEWY
QQVERGEKVRINIPEIKVADEEDRSVFIEVHEPLRSNGVKKNLVVPELSNWLQSHK
ISDAKARAALVLAQNGVPEIAPPQWETIAQ"
/complement (6830..7189)
/gene="VCA0059"
/complement (6830..7189)
/gene="VCA0059"
/notes="similar to GB:M13176 SP:P09461 GB:M28554 PID:150894
PID:150896; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="major outer membrane lipoprotein"
/protein_id="AAF95973.1"
/db_xref="GI:9657442"
/translation="MRASPARNCNVNSVHILLFDKYRTSITPTQVKEFSMNKMLIA
AAASVLLIAGCAGSDEATTATKMEISTQVSELNSQVAAALASKVDQAAEAAQAQEE
AARANERIDNTIAQSYTK"
/7386..7562
/gene="VCA0060"
/7386..7562
/gene="VCA0060"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95974.1"
/db_xref="GI:9657443"

```



```

Db      829  CCGCCACGACGACCGGATCGGTACACACCTTGGCGCGAGCGGTGGCTACCGC 888
Qy      712  ArgAsnThrArgTyrGlyGluThrAsnTrrTyrVallysAlaAspAsnLeuLeuAndGln 731
Db      889  TTCGACATTGGCCAGACCGAGTGGCGTGGTGGCGGCGAAGAACCTGACCGACCGAG 948
Qy      732  SerValTyrAlaHisSerPheLeuSerAsp---ThrProGlnMetGlyArgSerPhe 750
Db      949  ACCGTGGCTATGCCAGTTCGACCTTGGCGATACGACCGCGCGCGGCGGTACCGTG 1008
Qy      751  ThrGlyGlyVal 754
Db      1009  GAAGTCGGGTG 1020

RESULT 31
LOCUS   AE005425                10161 bp    DNA        linear    BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 44
of 290.
ACCESSION AE005425
VERSION   AE005425.1
KEYWORDS  GI:12516166
SOURCE    Escherichia coli O157:H7 EDL933
ORGANISM  Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 10161)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
MEDLINE
PUBMED 11206551
REFERENCE 2 (bases 1 to 10161)
AUTHORS  Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
TITLE     Submitted (22-OCT-2000) Laboratory of Genetics, University of
JOURNAL   Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
MEDLINE   Location/Qualifiers
PUBMED    1. 10161
FEATURES  source
           /organism="Escherichia coli O157:H7 EDL933"
           /mol_type="genomic DNA"
           /strain="EDL933"
           /serotype="O157:H7"
           /db_xref="taxon:155864"
           /note="enterohaemorrhagic"
           172..247
           /gene="asnV"
           /note="synonym: Z3149"
           172..247
           /gene="asnV"
           /product="tRNA-Asn"
           /note="anticodon: GTT"
           /function="RNA; tRNA"
           complement(303..1235)
           /gene="erfK"
           /note="synonym: Z3150"
           complement(303..1235)
           /gene="erfK"
           /function="orf; Unknown function"
           /note="Residues 1 to 310 of 310 are 99.03 pct identical to
           residues 1 to 310 of 310 from Escherichia coli K-12 Strain
           MG1655: B1990"
           /codon_start=1

gene     gene
tRNA     tRNA
gene     gene
CDS      CDS

```

```

/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG57051.1"
/db_xref="GI:12516167"
/translation="MRRVNIICSFALLFASHTSLAVTYPLPPEGSRLVQGSLTVTPD
HTNPGLTFAAYGQGLSNMLEANPGADVFLPKSGSOLTIPQQLILPVTVRKGLVWV
AMELYLPPDSNTVEFPFPIGIGAGRETFRNWTTVRKEQEAFTPTPTTRREYAK
RGESLPAPFVDPNPMGLYAIYIGRLYAIHTNANFGILGRVSGGIGIRLNDIKYL
FDNPVGTFRVQIIDQPKYTTPEPDGSKWLEVHEPLSRNRAEVESDRKVPLPVPFSLRA
FINGQEVNVRANAALQRSQMPVQISSGRQMF"
/gene="cobT"
/note="synonym: Z3151"
complement(1300..2379)
/gene="cobT"
/function="enzyme; Biosynthesis of cofactors, carriers:
Cobalamin"
/note="Residues 1 to 359 of 359 are 95.82 pct identical to
residues 1 to 359 of 359 from Escherichia coli K-12 Strain
MG1655: B1991"
/codon_start=1
/transl_table=11
/product="nicotinate-nucleotide dimethylbenzimidazole-P
phosphoribosyl transferase"
/protein_id="AAG57052.1"
/db_xref="GI:12516168"
/translation="MQTLANLLNTIPADPAAMGSAQRHIDGLLKPVSGSLGRLEALGV
QAAGPGLNGIPHGKAVLWMCADHGVWEEXVAISPKEVTAIOAENNRGTGTGVCVL
OLAAGANVHIDVGDIAPIPIGLINMRVAXSGNIASAPAMSRQAEKLLDVICYT
RELAKNGVTLFGVGLGMANTPAAATVSTTIGRDPERVVGIGLANPTDKLANKDVV
BRAITLQNPQDGVNVLAKYGGFDLVGMAGLMCAASCGPLVLLDGLFSLAALAAC
QMSPAIKFPLPSHUSAGKARIALSHGLEPFLNMDMRLEGGGAAALAMPIEAACA
LYNNNGELAAASKIVLPNGNTSDLNS"
/gene="cobs"
/note="synonym: Z3152"
complement(2391..3134)
/gene="cobs"
/function="enzyme; Biosynthesis of cofactors, carriers:
Cobalamin"
/note="Residues 1 to 247 of 247 are 99.59 pct identical to
residues 1 to 247 of 247 from Escherichia coli K-12 Strain
MG1655: B1992"
/codon_start=1
/transl_table=11
/product="cobalamin 5'-phosphate synthase"
/protein_id="AAG57053.1"
/db_xref="GI:12516169"
/translation="MSKLPWAMLSPTIRLPVPRMSQGLDFEHSRGIITPPLIGLLL
GAISGLVFWLQAWCGAPLAFSLVLVMTGPHLDGLADTCDFVSARSRRDMLLE
IMRDSGLTGHGLALFVYLAKILVLSLALRGEFILASLAAACAVSGTALLMYRH
RYAREGLGNVFIGIDGRQTCVTLGLAAIPAAVLPGMHGVAMVVTVAIFILGQL
LKRTIGGQTGDTLGAAILBELGLVFLALL"
/gene="cobU"
/note="synonym: Z3153"
complement(3131..3676)
/gene="cobU"
/function="enzyme; Biosynthesis of cofactors, carriers:
Cobalamin"
/note="Residues 1 to 181 of 181 are 100.00 pct identical
to residues 1 to 181 of 181 from Escherichia coli K-12
strain MG1655: B1993"
/codon_start=1
/transl_table=11
/product="cobinamide kinase/cobinamide phosphate
guanylyl transferase"
/protein_id="AAG57054.1"
/db_xref="GI:12516170"
/translation="WMILVTGARGSKSRHAELIGDSSQVLYTATSOILDMAARI
EHRQGRPEHRTVERWHLDELHAPINPEVLELCTTWTNLLDFYDGGDKDPDE
WDIQAEQAINAEIOTSLIARQCPKPVVLVTNVEGMGIVPESRLARFDIAGKNQ
QLAAANAEVNLVVSIGVKIK"

```


Db	8705	GAGCAGACATCCAGCGCAATTGCCAAC			8755	9683	CTTGTGATGGTCCAGCCCAACATAATCTCTGATGTTTATTTATTCGTGCTGCTCTAGTGCA	9742
Qy	343	GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnA:GAsnAspTyrHisHis			362	651	GlnAspAspGlnAsnAla	667
Db	8756	GTTCCAGTCCGATGATTCAGGCTCTGACCGTTCAT			8809	9743	AAAGATGACAGTAAACGGTACCGAAACAAATGTTCCGGCTGGCCCACTCTCAACTTTGCA	9802
Qy	363	AspGluLysAlaGlyAspAlaVal			380	668	LeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal	686
Db	8810	GATAAGCAATACACTCAGCGGTGCATGTCAGAGTCACCTTTCTGCTCCCTGCTTAATAT			8869	9803	GTTAATACAGAAATTC	9847
Qy	381	AlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrPheValGln			400	687	AlaGlnAsnLysLeuAla	701
Db	8870	GAACCTTGTTACC			8890	9848	GCACTCAATACCTGACAGACAAACGCTACCGTACGACATCAAACTATTCTTCGAGCA	9907
Qy	401	Tyr			416	701	shMetLeuAsnLeu	717
Db	8891	TACAACAACAGATAGGTTCAGCGAGTCCGCTGCGATGACCTCAAGCAAAATCTCTGACC			8950	9908	GGTTTAAATCAGCTATAGTGTGTTGATGGAATTTCTGTGACAAAACAACTGTTATCG	9967
Qy	417	LysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGln			436	717	Y	717
Db	8951	GCC			9001	9968	A	9968
Qy	437	Ala			445	RESULT 32		
Db	9002	AGTACAGTCTCACTATTCGCACAAATAGCTGCAATTCGCGATCACTGGACATGGACA			9061	AP002559	286857 bp	DNA linear BCT 07-MAR-2001
Qy	446	GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAsp			465	DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 10/20.	
Db	9062	ATGGAGTTCGC			9109	ACCESSION	AP002559	BA000007
Qy	466	ArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla			484	VERSION	AP002559.1	GI:13362012
Db	9110	CGAGTATATATACCGCAGCATATATAAGCATACCTCTCTTCCGACAGAGCTCGAGT			9169	KEYWORDS		
Qy	485	ArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeu			503	SOURCE	Escherichia coli O157:H7	
Db	9170	GATCAGAAATGTAACATCTACACCTGCGCTATTCAGTTTCGATTAACCTGGAGTTA			9229	ORGANISM	Escherichia coli O157:H7	
Qy	504	ThrAlaSerHisGlnGluArg			521	REFERENCE	1 (sites)	
Db	9230	CGCGTCCGTCGCGCAAGCTACGTATTTCCACACACTCTCCAGCTTTTATTCGACACA			9289	AUTHORS	Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H.	
Qy	522	LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer			541	JOURNAL	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak	
Db	9290	TCTGCGGCGGCGAGTGTACATAC			9343	MEDLINE	20198780	
Qy	542	AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrPheGlnTyrAsnLeuAlaLeu			561	PUBMED	10734505	
Db	9344	ATAAATTTTGAATAGGTGCGCATATATGTTAATGTTAATGTTGATTCGACAGCGCAGTT			9403	REFERENCE	2 (sites)	
Qy	562	TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro			581	AUTHORS	Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.	
Db	9404	TACTACTCAGAACTAAAGATTATAT			9460	TITLE	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655	
Qy	582	LysSerIleGluAspSerGluMetLysLeuValArgTyr			599	JOURNAL	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)	
Db	9461	TGCAATGTTAACAACCACTCTCCCTAGTAGTACTATTTATTCACATATTTGATCGG			9520	MEDLINE	20557356	
Qy	600	AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIle			619	PUBMED	11108008	
Db	9521	GCAAAACATGGGAGCTGGAA			9562	REFERENCE	3 (sites)	
Qy	620	GlyValSerGlyAspTyrValArgGlyArgLeu			630	AUTHORS	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H.	
Db	9563	TGGTTCCTCGCATATATCAGTGGCAATTTAATTCGTGCGCAATATGAACCTCAACA			9622	TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak	
Qy	631	LysAsnLeuProSerLeuProGlyArg			639	JOURNAL	Gene 258 (1-2), 127-139 (2000)	
Db	9623	TTAATAAACAACTAATACAGGTGAACCGCATAAACGCTATAGGCTGAAACATACCT			9682	MEDLINE	20564182	
Qy	640	GluAspAlaTyrClyAsnArgProPheIleAla			650	PUBMED	1111050	
						REFERENCE	4 (sites)	
						AUTHORS	Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.	
						TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12	

```

JOURNAL      DNA Res. 8 (1), 11-22 (2001)
MEDLINE      21156231
REFERENCE    11258796
AUTHORS      5 (bases 1 to 286857)
             Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
             Hayashi, T.
TITLE        Direct Submission
JOURNAL      Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
             Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
             Japan [E-mail: ken@gen-info.osaka-u.ac.jp,
             URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
             Fax: 81-6-6879-2047]
COMMENT      genome project.
FEATURES     Location/Qualifiers
             1..286857
             /organism="Escherichia coli O157:H7"
             /mol_type="genomic DNA"
             /strain="O157:H7"
             /sub_strain="RIMD 0509952"
             /db_xref="taxon:83334"
             63..1508
             /gene="ECs2545"
             63..1508
             /gene="ECs2545"
             /note="similar to YEBU ECOLI gi|1788140 percent identity
             96 in 481 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="putative rRNA methylase"
             /protein_id="BAB35968.1"
             /db_xref="GI:13362013"
             /translation="MLVQAHTVYFPDAFLTQMRAMPSTLSEDFLACQRLRRSIR
             VNTLTSVADFLRLTAQYGTWLTLPWCERGFWRDDEDLPLGSTAEHLUSGLFYIQ
             EASMLPVAAQFADQVRVMDVAAAPGSKTQTAAAMNNEGAILANFEASRVKVL
             HANIRCGTSTNLTDFGRVFGAALPEMFDAILLDAPCGEGVVRKDPDALKNWSP
             SNOETAAORELIDSAFHALRSGGTLVSTCTLNRENEVCLMLKETYPDAVGFPLP
             GLFPKANKLTFEGFLHVFPOIYDCGFFVARLKTQAI PALTPKPKYKGNFPFPL
             KREAAQIPQAAAGLAWDGNLRWQDKVLPVGVLEALIGKRESRLGKLAET
             HNKGYRQHEAVIALSPDENAFELTQBAEAEWYRGDRDVPQAAVPADVVIVTFQHQ
             PIGLAKRIGSRUNSYPRELVDRGKLFTSNA"
             1611..1862
             /gene="ECs2546"
             1611..1862
             /gene="ECs2546"
             /note="similar to B1836 ECOLI gi|1788141 percent identity
             100 in 83 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB35969.1"
             /db_xref="GI:13362014"
             /translation="WESTKMKTSVRIGAFEDIDDELHGHESPGDRTLTIPCKSDPDLCLM
             QLDAMDATSIPALLNGHESVLYTRYDQDSAWINRLA"
             1883..2158
             /gene="ECs2547"
             1883..2158
             /gene="ECs2547"
             /note="similar to B1837 ECOLI gi|1788142 percent identity
             98 in 91 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB35970.1"
             /db_xref="GI:13362015"
             /translation="MAGILSWLFFPRCKISPKLNGTAPHGDEMFAVLVFCVLDGGCE
             DIVDVYNTFQQCLYSMSDRIHGGCFPIEDFIDGFWRPAQYGFDF"
             complement (2159..2818)
             /gene="ECs2548"
             complement (2159..2818)
             /gene="ECs2548"
             /note="similar to PP8A ECOLI gi|1788143 percent identity
             98 in 219 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="phosphoprotein phosphatase 1"
             /protein_id="BAB35971.1"
             /db_xref="GI:13362016"
             /translation="MMKQAPVYQRIAGHQRHILWISGDIHGLCQLRRKLRWHCRFPD
             WDLITSGDVDRGQPSQRLCLQLLEQHWVRAVRGHEQMAMDALASQMSLWLNKGG
             WFLDALDNQOAKTALCKCOHLPFLFVHSRTGKHVTAHADYDDVYEWCKVDLH
             QVLWRSRLSERQKGGITGADHFWFGHTPLRHRYVDIGNLHYIDTCVAFGGELTFQL
             Q"
             complement (3211..3552)
             /gene="ECs2549"
             complement (3211..3552)
             /gene="ECs2549"
             /note="similar to B1839 ECOLI gi|1788144 percent identity
             98 in 113 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB35972.1"
             /db_xref="GI:13362017"
             /translation="MMKKSILAFLLTSSAALAAPOVITVSRFEYKDKWAFNREEV
             MLTCRPNALYVINPTLYQVPLNDVAQKVASGKTKAQPISVIQIDDPNPGKMSL
             APFIERAEXLC"
             complement (3565..4437)
             /gene="ECs2550"
             complement (3565..4437)
             /gene="ECs2550"
             /note="similar to B1840 ECOLI gi|1788145 percent identity
             98 in 290 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="putative resistance protein"
             /protein_id="BAB35973.1"
             /db_xref="GI:13362018"
             /translation="MLAFTWIALRFHFTSLMLVFGFMYGAWLAPLTLRLTLKRF
             RLQCHAAVNSLISATAMLVQGLMGTKGVKDFSPNIWQAVLQTFQGVWLQVILAL
             VTILVALMQPNRPRLIFMLTQAQFILLAGSHATLNEGVTAKIHQTNHAIHLICAA
             WFGGLLPVLWCMOLIKGRWQHQAIOALMRFSWCGHFAVIGVLASGVNALLITGPPT
             LTYWQGLLILKAILVMIMVVALANRYVLVPRMQDEDEAAFPWFVMTKLEWAGAV
             VLLIISLLATLEFP"
             complement (4441..4815)
             /gene="ECs2551"
             complement (4441..4815)
             /gene="ECs2551"
             /note="similar to B1841 ECOLI gi|1788146 percent identity
             99 in 124 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="BAB35974.1"
             /db_xref="GI:13362019"
             /translation="MASTABSLYALAILTSLVTPSVWAHAILTHQYPRANAQVTA
             PQAITLNFSGVETGFGCAKITGPKNENIKTLPAKNEQDKRLIPLADSLKGYTYT
             VDHVSVSDGKHGKTHGYTFSVK"
             4954..5184
             /gene="ECs2552"
             4954..5184
             /gene="ECs2552"
             /note="similar to HOLE ECOLI gi|1788147 percent identity
             100 in 76 aa (Conserved in E.coli K-12)"
             /codon_start=1
             /evidence=not experimental
             /transl_table=11
             /product="DNA polymerase III theta subunit"
             /protein_id="BAB35975.1"
             /db_xref="GI:13362020"

```



```

Db      219135 TGGGTTTCTCGCCATATATCAGTGGCAATTTAATTCGTCGCAATATGAAATCTCAACA 219194
Qy      631 -----LysAsnLeuProSerLeuProGlyArg----- 639
Db      219195 TTAATAACAACTAATACAGGTGAACACCGGATAACCGACGTATATAGGCTGAAACATACCT 219254
Qy      640 -----GluAspAlaTyrGlyAsnArgProPheIleAla 650
Db      219255 CTTGTGATGGCTCAGGCCAACATAATCTCTGATGTTTTATTCGTGCTGCCTCTAGTGCA 219314
Qy      651 GlnAspAspGlnAsnAla-----ProArgValProAla---AlaArgLeuGlyPheHis 667
Db      219315 AAAGATGACAGTAACAGGTACCGGAAACAAATGTCGCGGCTGGCGACCTCTCAACTTTCGA 219374
Qy      668 LeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal---Phe 686
Db      219375 GTAAATCAGAAATC-----GGTAAACGAGGATCAGTACCGGATTAACCTG 219419
Qy      687 AlaGlnAsnLysLeuAla-----ArgTyrGluThr-----ArgThrProGly-Hi 701
Db      219420 GCACTCAATACCTTGACAGACAAACGCTACCGTACAGCACATGAACCTATTCTTCGCGACA 219479
Qy      701 ShiSerLeuAsnLeu-----GlyAlaAsnTyrArgArgAsnThrArgTyrGly 717
Db      219480 GGTTTTAATCAGCTATATAGGTTTGTATGGAATTTCTGATGACAAAAACAATCGTTATGG 219539
Qy      717 Y 717
Db      219540 A 219540

RESULT 33
LOCUS   AE016762
DEFINITION Escherichia coli CFT073 section 8 of 18 of the complete genome.
ACCESSION AE016762 AE014075
VERSION   AE016762.1 GI:26108544
KEYWORDS
SOURCE   Escherichia coli CFT073
ORGANISM Escherichia coli CFT073
REFERENCE 1 (bases 1 to 311143)
AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
12471157
REFERENCE 2 (bases 1 to 311143)
AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Direct Submission
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. 311143
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement(206..952)
/locus_tag="c2289"
complement(206..952)
/function="putative transport"
/notes="Residues 103 to 248 of 248 are 97.26 pct identical
to residues 1 to 146 of 146 from MG1655 : b1874"
/codon_start=1
/transl_table=11

gene
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement(206..952)
/locus_tag="c2289"
complement(206..952)
/function="putative transport"
/notes="Residues 103 to 248 of 248 are 97.26 pct identical
to residues 1 to 146 of 146 from MG1655 : b1874"
/codon_start=1
/transl_table=11

CDS
1..311143
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement(206..952)
/locus_tag="c2289"
complement(206..952)
/function="putative transport"
/notes="Residues 103 to 248 of 248 are 97.26 pct identical
to residues 1 to 146 of 146 from MG1655 : b1874"
/codon_start=1
/transl_table=11

```

```

/product="Putative conserved protein"
/protein_id="AA080748.1"
/db_xref="GI:26108545"
/translation="WALLIEICYSMECALTAQONCADRVLCAPKKEGLTSLGLVIX
SVORVITPVHPIIRPRGDFCYSDGEFAALLEVYTVRELGFPGLVTVGLVDGND
MPREKMAAGAPLAVTFHRAFDPCANPINTLNLTIELGIAKVLTVGGKSLDQGLSK
IMELIARDAFLINAGAGVAKENLHHFDAGVLEVHSSAGAWQASPMRYRNGLSMSS
DAHADYSRYVVDGAAVAMKGIIEIHOAK"
complement(966..1538)
/gene="yecM"
/locus_tag="c2290"
complement(966..1538)
/gene="yecM"
/locus_tag="c2290"
/notes="Escherichia coli K-12 ortholog: b1875; Escherichia
coli O157:H7 ortholog: z2928"
/codon_start=1
/transl_table=11
/product="Protein yecM"
/protein_id="AA080749.1"
/db_xref="GI:26108546"
/translation="MIMANWOSIDELQDIASDLPRFTHALDELRLGLDITPLTADH
ISLRCHONVTAERWRGPEOCGETLLSENWINGRPICLFELHEPVQVAHQWSIVELPW
PGEKRYPEHGEHIEIIVLPGDPETLNARALALLSDGLSLFGIIVKTSPPKGEHERLP
NPTLAVTDGKTIIFKHPWSEIEIVASEOSA"
1748..3481
/gene="argS"
/locus_tag="c2291"
1748..3481
/gene="argS"
/locus_tag="c2291"
/EC_number="6.1.1.19"
/function="enzyme; Aminoacyl tRNA synthetases, tRNA
modification"
/notes="Escherichia coli K-12 ortholog: b1876; Escherichia
coli O157:H7 ortholog: z2929"
/codon_start=1
/transl_table=11
/product="Arginyl-tRNA synthetase"
/protein_id="AA080750.1"
/db_xref="GI:26108547"
/translation="MNIQALLSEKVRQAMTAAAGADCEPQVRSQAKVQFGDYQANGM
MAVAKGMAPRLQAEVLVLDLNGHASKVEITAGPGFINIFLDPFALAEHQOALAS
DRUGVATPEKQITVDVYSAPNVAKEHMGVHLRSTIIGDAVRTLFLGHKVRVANHVG
DWGTQFGLMIALEKQOENAGEMELADLEGFYRDAKHVDEDEEPARARNYVVKLQ
SGDEYFRMRKLVDTMTQNTIYDRLNVLTFDDVMGSLYVPMPLPGIVADLKAQG
LAVESGATVVFLEDFKNKEGEPMTIIOKDGGLYTTTDDIACAKRYETLHADRVIL
YVIDSQHQHLMQAWAIVKAGVPSVPLEHMFGLMGKDGKPFKTRAGGTGVLKAD
LLDEALERARLVAKRNPMPADELEKLANAVGAVYADLSKNRTTDTYTFDWMML
AFEGNTAPYQYAYTVLVSVPKAEINEEQALAAPVIRREDREAQLAARLQFETILT
VVAEGTPTVMCAIYLDLAGLFGSGFVHCPILSAENEVRNRLKLAQLTAKTLKGL
DTLGIETVERM"
complement(3534..3926)
/gene="flhE"
/locus_tag="c2292"
complement(3534..3926)
/gene="flhE"
/locus_tag="c2292"
/function="structural component; Surface structures"
/notes="Escherichia coli K-12 ortholog: b1878; Escherichia
coli O157:H7 ortholog: z2931"
/codon_start=1
/transl_table=11
/product="Flagellar protein flhE precursor"
/protein_id="AA080751.1"
/db_xref="GI:26108548"
/translation="MRALLAILFPLLVOAGEMQWQASVGVTLNHRGESSAPLS
TRQPAAGMTLVAWRYQLIPTPAGLRVLCRSQRCVLEGGSGTTVAFGIAABPL
RTIWEVPGGRLIPLPKIQIRNEVINYR"
complement(3926..6079)
/gene="flhA"
/locus_tag="c2293"
complement(3926..6079)

```



```

/gene="flhA"
/locus_tag="c2293"
/function="structural component; Surface structures"
/notes="Escherichia coli K-12 ortholog: b1879; Escherichia coli O157:H7 ortholog: z2932"
/codon_start=1
/transl_table=11
/product="Flagellar biosynthesis protein flhA"
/protein_id="AAN80752.1"
/db_xref="GI:26108549"
/translation="MDSALYNLLIFCRKPMWILLTRNPMNSLAAMLRLPANLKSSTOW
QLAGPILLLFLLSMWLPFLAFILDLLEFTNLISIMVLVAMFTQRTLEFAFPPI
LFTLLRLALNVASTRILMEGTCAAAGKVEAFGLVGGNFAIGIVFVILVI
INFWITGAGRIAEVGFARFLVDGPGQMAIDADLNAGLIGEDAEKRRREVTEAD
FYGMGASKFVRGDAIGILIMVNVGGLLVGLQHGSMGHAESYTLATIGDGL
VAQIPALVISTAAGVIVRVSTDDQVGEOMVNLFSNFSMLLSAVALGILGLVPGMP
NVLFLFTAGLGLAWLRGRKQPAEPKPFVPMANNIVVEATWVNDWDSIGMEV
GLRIPMDVFDQDQGLFIRSRKFAQMGFLFPVPHIRHMDLQPARYRLMGV
ETGSDAYFGRVILNPGTAAGTLPFGALNATWIESALKKEQAQYQYTVVE
ASTVATHLNLHISQAAELFGRAQQLDRVAQEMPKLTEDLPVGVVTLTLHKVL
QNLDEKVPIDRMRTILETASHAPIQSDPHLTAVRVALGRAITQOMFPKDEHVH
IGLDTPLERLLQALQGGGLPEPLADRLIAQTOEALSQEMLGAPPVLLVNHALRPL
LSRFLRRSPOLAVLSNLELSDNRHIRMATIGK"
/complement(5997..7145)
/gene="flhB"
/locus_tag="c2294"
/complement(5997..7145)
/gene="flhB"
/locus_tag="c2294"
/function="structural component"
/notes="Escherichia coli K-12 ortholog: b1880; Escherichia coli O157:H7 ortholog: z2934"
/codon_start=1
/transl_table=11
/product="Flagellar biosynthetic protein flhB"
/protein_id="AAN80753.1"
/db_xref="GI:26108550"
/translation="MSDESDDKTEAPTHLEKABEEQIPRSRELTSLILLVGVSV
INFGVSLARRISGMLSAGLHFDHIIINDPNILQIILLIREAMLLPLISGVLV
ATISPMVGLGVFSGLPKPSKLNPLPGIKRMFSAGTGAELKAILKTLIVGSVTG
FLIWHMPQMLMAESPTIATMGNAVDLGLCALLVGLVPMVGFDPVFFQIFSHLKK
LRMSQDIDRFQSGSDPHVRIRQMAAARRRMADVPKADVIVNNPHTYSVAL
QYDNEKMSAPKVKAGLVLRIRIIGAENVPTLEAPILARLYRHAIEIGQIPGQ
LYAAVEVLAVWQLKRWELAGQRPVQPTPLPVEALDFINEKTHE"
/1718..7324
/locus_tag="c2295"
/1718..7324
/locus_tag="c2295"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN80754.1"
/db_xref="GI:26108551"
/translation="MMPERKAFNGLSKGNRRDLGLLPQHIRHGMDFVGYGVNALSNI
EDLM"
/gene="cheZ"
/locus_tag="c2296"
/complement(7354..8028)
/gene="cheZ"
/locus_tag="c2296"
/complement(7354..8028)
/gene="cheZ"
/locus_tag="c2296"
/function="regulator; Chemotaxis and mobility"
/notes="Escherichia coli K-12 ortholog: b1881; Escherichia coli O157:H7 ortholog: z2935"
/codon_start=1
/transl_table=11
/product="Chemotaxis protein cheZ"
/protein_id="AAN80755.1"
/db_xref="GI:26108552"
/translation="MKNWACEDATMMQPSIKPADEHSGADITARIQSITRMLRSLRE
LGLDQIAIAEAAEAPIDARLRYVQMTAQAAERALSVEASQPHQDWEKSALKTQ
RWDWFPADPIDLADARELVTDTRQFLADYPAHTSFNACOLLIMMAQDFDLTGQVIK
RWDVLTQIERQLLWVLENIPEQESRPKRENSLLNGEPQVDTSKAGVVASQDQVDDL

```

```

LDSLGP"
complement(8009..8398)
/gene="cheY"
/locus_tag="c2297"
complement(8009..8398)
/gene="cheY"
/locus_tag="c2297"
/function="regulator; Chemotaxis and mobility"
/notes="Escherichia coli K-12 ortholog: b1882; Escherichia coli O157:H7 ortholog: z2936"
/codon_start=1
/transl_table=11
/product="Chemotaxis protein cheY"

Alignment Scores:
Pred. No.: 0.000215 Length: 311143
Score: 265.00 Matches: 176
Percent Similarity: 38.16% Conservative: 122
Best Local Similarity: 22.54% Mismatches: 338
Query Match: 6.57% Indels: 145
DB: 1 Gaps: 40

US-09-936-377-2 (1-758) x AE016762 (1-311143)
Qy 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSer 30
Db 212499 CTGTCAATATGCAATTCCTGGGATAGTTCTTACTGCAACCGCACTTCTTCAGAAACAAA 212558
Qy 31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
Db 212559 ATCAGCAACGAGAGAGCGCTCGTCGTGACCAACCAATCGT-----TCGGCAAGCAAC 212609
Qy 51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
Db 212610 CTTTGGGAAAGCCCGGCGACTATACAGTTATTGACCAACAAACATTCGACAGCTCCACC 212669
Qy 71 ---ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGly 89
Db 212670 AATCCCTCCATACCGGATAATTGACAGACATCCCGGAGTAGAGATAACAGCAACATCC 212729
Qy 90 ---GlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeu 108
Db 212730 TTGGCAGCCGCTAAACAAATCCCGATTTCGTGGGAGCATCCCTCCGTTTAAATTC 212789
Qy 109 AsnHisGlyGlu---ThrGlyAspMetAlaAspPheSerProAspHisAlaIleMet 127
Db 212790 ATTGATGTCAGGAGGTAACTTATCAGCGCGCGGAGATTAATATGCTGTGGGACTGTG 212849
Qy 128 ValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyr 147
Db 212850 ATAGATGAGTCTGCGTGGAGCGTGTGAGGTAGTGAAGGTCCATATTCGTTACTGTAC 212909
Qy 148 SerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMet 167
Db 212910 GGTTCACAGGCAATGGCGGTATTGTTAACTTCATACCAAAAAGGGAGGTGCAAACT 212969
Qy 168 ProGluAsnGlyValSerGlyGluLeuArgLeuSerSerGlyAsnLeuGluLys 187
Db 212970 ---GCATCTGAGTGTGAAGAGCTGTTTATAATTCACACAGCGCGCTGGAGATCA 213026
Qy 188 LeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGly 207
Db 213027 ATCCGGTCCAGGG---AGCATCGGT-----GGATTTCATTATCGCATCAACGGT 213074
Qy 208 LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro 227
Db 213075 AGTTATTCTCATCAGGCAATCGTGATACCGCGGAT-----GGACGTCTGCGG 213122
Qy 228 AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg 247
Db 213123 AATACC-----AACTATCTGTAACAAVAGTCAGGGTGTATGGTTGGTTATACTCCGA 213176
Qy 248 PheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHis 267

```

```

Db 213177 AACATCGT-----TTGGCCTCGCTGATCGCTACAGACTCGCGACGCAACTTAC 213230
QY 268 GluTyrAspAspCysHisAlaAspIleTyrPheGlnLysSerLeuIleAsnLysArgTyr 287
Db 213231 ---TATGAGGATCCAGACGGAAGCTAT-----GAGGCATTT 213263
QY 288 LeuGlnLeuTyrProHisLeuLeuThrGluGlu-----AspValAspTyr 302
Db 213264 AGTGTCAAAATACCTTAACCTTGAACGAGAGAAAGTTGGGGTATTCTATGACACAGACGTG 213323
QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAlaHis 322
Db 213324 GACGGTACTATCTAAAA-----AAATTCATTTCCGACGCGTAT 213362
QY 323 AsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342
Db 213363 GAGCAGACCATCCAGCGCAATTTGCCAAC-----GAAGTAAACACGACACAGCCT 213413
QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHis 362
Db 213414 GTTCCAGTCCGATGATTGAGCTCTGACCGTTTCAT---AACAGACTGAC---ACCCAT 213467
QY 363 AspGluLysAlaGlyAspAlaVal-----GluAsnPheAsnAsnGlnThrGlnAsn 380
Db 213468 GATAAGCAATACACTCAGCGGTCACTTGCGAGAGTCACTTTTCGCTGCTGCTGCTTAATAAT 213527
QY 381 AlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGln 400
Db 213528 GAACCTGTGTACC-----GGTGCACAG 213548
QY 401 Tyr-----LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaVal 416
Db 213549 TACAAAACAGACAGCGGTGACGCAAGTCCGCTGCGTACGACCTCAAGCAAACTCTCTGACC 213608
QY 417 LysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGln 436
Db 213609 GGC-----TTCAATTAATGAAGAAACAGCACTCGCTCTCTATATGATGATCAGAGCA 213659
QY 437 Ala-----AsnTrp-----AspAsnPheThrLeuGlu 445
Db 213660 AGTACAGTCTCACTATTGCGACAAATGACTGCGGATTCGCGGATTCAGTGCAGATGGACA 213719
QY 446 GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAsp 465
Db 213720 ATGGGAGTTTCG-----CAATACTGGCTTCTTCAAGTGTGACCGCGTGT-----GAC 213767
QY 466 ArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla--- 484
Db 213768 GGAGTATCATATACCGCAGCGCATTTAAGCGATACCTCTTTCGCGAGAGTCTCGAGT 213827
QY 485 ---ArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSerLeu 503
Db 213828 GATCAGAAATGGTAACATCTACAAGCTCGCCTATTGAGTTCGATTAACCTGGAGTTA 213887
QY 504 ThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuTyrAlaHisGly 521
Db 213888 CGCGCTCGCTTCGCGCAGGCTACGTATTTCACACACTCTCCAGCTTTTATGACAGACA 213947
QY 522 LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer 541
Db 213948 TCTCGCGCGCGCATGTCACATAC-----GAAATCTCTGATCTTAAAGCTGAACACTCC 214001
QY 542 AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeu 561
Db 214002 AATACTTTGAATTAGTGCAGATATAATGTTATCATGCTGCTGATGACAGCGCAGTT 214061
QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
Db 214062 TACTACTCAGAGCTAAAGATTATATT---GCAAGTCTGATCTGTGATGGCAGTATGTT 214118
QY 582 LysSerIleGluAspSerGluMetLysLeuValArgTyr-----AsnGlnSerGly 599

```

```

Db 214119 TGCATGGTAAACACCAACTCTCCCGTAGCTACTATTATTATGACAATATTGATCGG 214178
QY 600 AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIle 619
Db 214179 GCAAAACATGGGACTGGAA-----ATAAGCGCGAATATATATGCG 214220
QY 620 GlyValSerGlyAspTyrValArgGlyArgLeu----- 630
Db 214221 TGGGTTTTCTGCCATATATCATGTCGCAATTTAATTCGTCGCAATATGAAACTTCAACA 214280
QY 631 -----LysAsnLeuProSerLeuProGlyArg----- 639
Db 214281 TTAATAAACCACTAATAACAGGAGAACCGGATAAACCGACGTATAGGGCTGAAACATACT 214340
QY 640 -----GluAspAlaTyrGlyAsnArgProPheIleAla 650
Db 214341 CTGTGATGGTTCAGGCCAACATAATCTCTGATGTTTTTATTTCGTCTGCTCTAGTGCA 214400
QY 651 GlnAspAspGlnAsnAla-----ProArgValProAla---AlaArgLeuGlyPheHis 667
Db 214401 AAAGATGACAGTAAACGCTACCGAAACAAATGTTCCGGGCTGGGCCACTCTCAACTTTGCA 214460
QY 668 LeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal---Phe 686
Db 214461 GTAAATACAGAAATC-----GGTACAGAGATCAGTACCGGATTAACCTTA 214505
QY 687 AlaGlnAsnLysLeuAla-----ArgTyrGluThr-----ArgThrProGly-Hi 701
Db 214506 GCATCAATTAACCTGACAGACAAACGCTACCGTACAGACATGAAACTATTCTCGCAGCA 214565
QY 701 sHisMetLeuAsnLeu-----GlyAlaAsnTyrArgArgAsnThrArgTyrGly 717
Db 214566 GGTTTTAATCAGCTATAGGTTTTTGTATGCAATTTCTGTGACAAAACAAATCGTTATGG 214625
QY 717 Y 717
Db 214626 A 214626

```

```

RESULT 34
AF337562 2589 bp DNA linear BCT 29-MAY-2001
LOCUS
DEFINITION Escherichia coli clone RDA1:ex unknown gene.
ACCESSION AF337562
VERSION AF337562.1 GI:14211709
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 2589)
Allen,N.L., Hilton,A.C., Betts,R. and Penn,C.W.
Use of representational difference analysis to identify Escherichia
coli O157-specific DNA sequences
FEMS Microbiol. Lett. 197 (2), 195-201 (2001)
JOURNAL
MEDLINE 21213883
PUBMED 11313134
REFERENCE
2 (bases 1 to 2589)
Allen,N.L. and Penn,C.W.
Direct Submission
AUTHORS
TITLE
Submitted (18-JAN-2001) School of Biosciences, University of
Birmingham, Birmingham B15 2TT, UK
JOURNAL
FEATURES
Location/Qualifiers
1..2589
/organism="Escherichia coli"
/mol_type="genomic DNA"
/serotype="O157"
/db_xref="taxon:562"
/clone="RDA1:ex"
471..>2589
/notes="similar to TonB dependent outer membrane proteins"
/codon_start=1
/transl_table=11
/product="unknown"

```

```

CDS

```

```

/protein_id="AAK57482.1"
/db_xref="GI:14211710"
/translation="WRDEWLYNIPCRIVILSTLSLCISGIYSTATATSETSKISNEET
LVVITNRSASNLWESPATIQVDDQTLQNSINASADNLDQIPGVETITDNLAKRQKI
RIRGEASSRLVLDQEVITYORAGNNGVGLLIDDESALERVEVVKVGYSLYGSQAI
GGIVNFTIKKGGDLKQAGVAVYNSATAGWBEESTAVQSGISGDFYRINGSYDQGNR
DTPDGLRPNVYNNNSGVLGYSNGHFRGLSLDRYLATQTYVEDPDGSEYFASVK
IPKLERKGVFVDVDDGDKLKHIFDAYEOTIORFANEVKTTPVPSPMIQALTIV
HNKTHDKOYTOAVTLOSHFSLPANNELVTGAOKQDVRSQSGMSSKSLGTFIN
KETRHSYESSQSVLSLPQANDWQFADHWNTMGVQRYWISSKLITRGDGVSYTAGII
SDTSLAARESSQEMVSTSLRYSGFDNLLELRAAFQGVFFTLSQLFQTSAGSIVT
YGNPDJKAHNSNFELGARYNGNWLIDSAVYSBAKDIYASLDICDGSIVCNNGTNS
RSYYDYDDNNKWTGLIEISAEYNGWVPSYISGNLIRROYETSTLTKTGTGPAIN
GRIGLGLTUVGMAQNTISDVFIRAAASSAKDNGSTETNVPGWATLNFVAVTEFGNEDQ
YRIINLALNLTDKRYRTAHEIPA"
BASE COUNT      781 a   571 c   561 g   675 t   1 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      5,27e-07      Length:      2589
Score:          264.50        Matches:      169
Percent Similarity: 38.16%    Conservative: 118
Best Local Similarity: 22.47% Mismatches:      328
Query Match:     6.55%       Indels:        137
DB:              1          Gaps:           38

```

US-09-936-377-2 (1-758) x AF337562 (1-2589)

```

QY 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSer 30
DB 525 CTGTCATTATGCAATTTCTGGATAGTTTCTACTGCAACCGCACTCTTCAGAAACAAA 584

QY 31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
DB 585 ATCAGCAACGAGAGACGCTCGTCTGACCAACATCGT-----TCGGCAAGCAAC 635

QY 51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
DB 636 CTTTGGGAAGCCGGGACATATACAGGTATTATGACCACCAACATGTGACAGACTCCACC 695

QY 71 ---ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnThrGly 89
DB 696 AATGCCTCCATAGCCGATAATTTTGCAGGACATCCCGGAGTAGAGATAACAGCAACTCC 755

QY 90 ---GlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeu 108
DB 756 TTGCGAGGCGTAAACAAATCCGCAATTCGGGGAAGACATCCTCCCGTGTATTAAATCTC 815

QY 109 AsnHisGlyGlu---ThrGlyAspMetAlaAspPheSerProAspHisAlaIleMet 127
DB 816 ATTGATCGTCAGGAGGTAATTCATCAGCGCGCGGAGAGATAATTATGCTGTGGGACGTGTG 875

QY 128 ValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuThr 147
DB 876 ATAGATGAGTCTGCGCTGCGGCGGTGTGAGGTAGTGAAGGTCCATATTCGCTGCTGATC 935

QY 148 SerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGlyLysMet 167
DB 936 GGITTCAGGCAATTTGGCGGTATTGTTAATTCATCACCACCAAGGAGGTGACAACTT 995

QY 168 ProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLys 187
DB 996 ---GCATCTGAGTGTGAAGAGCTGTTTATAATTCGCAACACAGCGCTGGGAGAGATCA 1052

QY 188 LeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGly 207
DB 1053 ATCCGGTCCAGGGG---AGCATCGGT-----GGATTGATATTCGATCAACCGGT 1100

QY 208 LeuThrArgLysSerGlyAspThrAlaValProArgThrArgAsnLeuLysArgLeuPro 227
DB 1101 AGTTATTCTGATCAGGGCAATCGTGATACGCGGAT-----GGACGTCTGCGCG 1148

QY 228 AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyThrArgLysArg 247

```

```

Db 1149 AATACC-----AACTATCGTAACAATAGTCAGGGTGTATGGTTGGTTATAACTCCGA 1202
QY 248 PheThrArgArgThrThrSerAspArgArgAspGlnThrGlyLeuProAlaHisSerHis 267
Db 1203 AACCATCGT-----TTTGGCCTCTCGCTTGTATCGGTACAGCTCGCGACACAACTTAC 1256
QY 268 GluThrAspAspCysHisAlaAspIleIleThrGlnLysSerLeuIleAsnLysArgThr 287
Db 1257 ---TATGAGATCCAGACGGAAGTTAT-----GAGGCATTTT 1289
QY 288 LeuGlnLeuThrProHisLeuLeuThrGluGlu-----AspValAspThr 302
Db 1290 AGTGTCAAAATACCTAACTTGAACGAGACAAAAGTTGGGGTATTCTATGACACAGACGTG 1349
QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAlaHis 322
Db 1350 GACGTCGACTATCTAAAA-----AAATTCATTTCCGACGCGTAT 1388
QY 323 AsnGlyLysProThrIleAspLeuArgAsnLysArgThrGluLeuArgAlaGluThrLys 342
Db 1389 GAGCAGACCATCCAGCGCAATTTGCCAAC-----GAAGTAAAAACGACACAGCCT 1439
QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspThrHisHis 362
Db 1440 GTTCCAGTCCGATGATTCAGGCTCTGACCGTTCAT---AACAAAGACTGAC---ACCAT 1493
QY 363 AspGlyLysAlaGlyAspAlaVal-----GluAsnPhePheAsnAsnGlnThrGlnAsn 380
Db 1494 GATAAGCAATACACTCAGCGGTGCACATTCGAGTGCATCTTTTCGCTCGCTGCTAATAAT 1553
QY 391 AlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerThrGlyValGln 400
Db 1554 GAACCTGTGTACC-----GGTGCACAG 1574
QY 401 Tyr-----LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaVal 416
Db 1575 TACAACAGACAGAGGTGAGCCAAAGTCCGGTGGCATGACCTCAAGCAATCTCTGACC 1634
QY 417 LysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGln 436
Db 1635 GGC-----TTCAATTAATAGAAACACAGAACTCGCTCTTATATAGTCAGAGCAA 1685
QY 437 Ala-----AsnTrp-----AspAsnPheThrLeuGlu 445
Db 1686 AGTACAGTCTCATTCGCAAAATAGCTGGCAATTCGCCGATCACTGGACATGACA 1745
QY 446 GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAsp 465
Db 1746 ATGGGAGTTGCG-----CAATACTGGCTTTCTTCAAAAGTTGACGCGTGGT-----GAC 1793
QY 466 ArgGluAsnThrThrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla--- 484
Db 1794 GGAGTATCATATACCGCAGGCAATTAAGCATACCTCTTTCGACAGAGATGTCGAGT 1853
QY 485 ---ArgSerPheAlaLeuSerGlyAsnThrPheThrProGlnHisLysLeuSerLeu 503
Db 1854 GATCAGCAATAGTAAACATCTACAGCTCGCTATTTCAGGTTCGATAAATCGAGTTA 1913
QY 504 ThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuThrAlaHisGly 521
Db 1914 CGCGTGGCTTCGCGCAAGGCTACGTATTTCCACACTCTCCAGCTTTTTCGACACA 1973
QY 522 LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer 541
Db 1974 TCTGGCGCGCGAGTGTCAATAC-----GMAATCTCTGATCTTAAGGCTGAACACTCC 2027
QY 542 AsnAsnIleGluLeuAlaLeuGlyThrGluGlyAspArgThrGlnThrAsnLeuAlaLeu 561
Db 2028 AATACTTTGAATAGGTGCAGATATAATGGAATAGTGGTGTGATGAGTGCAGCGAGTT 2087
QY 562 TyrArgAsnArgPheGlyAsnThrIleThrAlaGlnThrLeuAsnAspGlyArgGlyPro 581

```

Direct Submission
Submitted (04-FEB-2003) Department of Molecular Biology and
Biochemistry, Simon Fraser University, 8888 University Dr.,
Burnaby, British Columbia V5A 1S6, Canada

This represents the February 3, 2003 version of the continually
updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation,
from PseudoCAP (see <http://www.pseudomonas.com> for latest updates
and links to alternate annotations). PseudoCAP is coordinated by
Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
E.W. Hancock (University of British Columbia, Canada). We welcome
submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the
accuracy of the protein name.
Class 1: Function experimentally demonstrated in P. aeruginosa.
Class 2: Function of highly similar gene experimentally
demonstrated in another organism (and gene context consistent
in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino
acid motif, structural feature or limited sequence similarity
to an experimentally studied gene.
Class 4: Homologs of previously reported genes of unknown function,
or no similarity to any previously reported sequences.

Location/Qualifiers
1. .16841
/organism="Pseudomonas aeruginosa PAO1"
/mol_type="genomic DNA"
/strain="PAO1"
/db_xref="taxon:208964"
74. .1009
/locus_tag="PA3264"
/feature="synonym: yocS"
74. .1009
/locus_tag="PA3264"
/feature="Protein name confidence: Class 3 (function proposed
based on presence of conserved amino acid motif,
structural feature or limited sequence similarity to an
experimentally studied gene)"
/codon_start=1
/transl_table=11
/product="probable transporter"
/protein_id="AAG06652.1"
/db_xref="GI:9949389"
/translation="MRALTAISRFVSAPFVLLFAILAFPPANFLPLTAIAIPL
GLVFMGMLTKGEDFREVAPRIYLVGLQFVFMGGLWLLCRLQLPAIANGVA
ILVGCPCGTASNTWNLSEGDVALSVATSVTLLAPLTPALWLLSASAWLPFSA
AFSLTQVVLPIAGLLAQRLLGRTQKVAEPLVSFVSIWIIIAAVVAASQARI
ASGLGLIMAVMLHNGFGLLLGYLTGKLTGMPLAQKRALAEVNGSLGALANAKL
FSPLAAPSAISFVWNLSGSLAALFRLLDTPR"
1149. .1463
/locus_tag="PA3265"
1149. .1463
/locus_tag="PA3265"
/feature="Protein name confidence: Class 3 (function proposed
based on presence of conserved amino acid motif,
structural feature or limited sequence similarity to an
experimentally studied gene)"
/codon_start=1
/transl_table=11
/product="probable transporter"
/protein_id="AAG06653.1"
/db_xref="GI:9949390"
/translation="MSWILIFLAGLFEVGVAVGLKTEGFSKPLPTVLTALAMLVSGI
LLGLAMKHLPLGTAYAVMTGVGAVGTIAGIVLFGESMALLRLASVALIVCGVLGLKL
SH"
1678. .1887
/feature="capB"
/locus_tag="PA3266"
/feature="synonym: cspa"
1678. .1887
/feature="capB"

Db 5559 GGGACCGCGCAACGGCTGGC-----GTGGCGCTGCTGTTATTCGGGGTG----- 5512
 QY 211 LysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerPro 230
 Db 5511 AAGGGCGCGCATAC----- 5497
 QY 231 ArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArg 250
 Db 5496 ---CGGACGGCAACACGACACCATATCGACAGCTGCTGCTCAACCCATTGGCAG 5440
 QY 251 ArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAsp 270
 Db 5439 CTCACCGACGCGCAGC-----TTGGCGGCCAACCTTCACCTACTAC--- 5398
 QY 271 AspCysHisAlaAspIleTyrGlnLysSerLeuLeuAsnLysArgTyrLeuGlnLeu 290
 Db 5397 GAGCGCTACGCCGATATG----- 5380
 QY 291 TyrProHisLeuLeuThrGluAspValAspTyrAspAsnProGlyLeuSerCysGly 310
 Db 5379 ---CGGGCGGCTGACCCAGCGCAGTACGAC---GACGATCCCTTCAGTCCGTA--- 5329
 QY 311 PheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAspLeu 330
 Db 5328 -----CGCGACTGGGACAATTTC 5311
 QY 331 ArgAsnLysArg----- 334
 Db 5310 CGTGGTGGCGGCAAGACTTCTCGTGAAGTACACCGCCGCGAGTGCAGACCTCACCCAG 5251
 QY 335 TyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHis 354
 Db 5250 TTCCAGGTGCTCACTACTACAGCAGCAGTTCCTCCGCGCAGCAGCATCGCGCGGCAAC 5191
 QY 355 LeuAsnArg-----AsnAspTyrHisAspAspGluLysAlaGlyAspAla 369
 Db 5190 CTCAGGACCATCACTCGTACCGCGCGGACTACCATGTGTTCCGGTGGAGCGCGGTC 5131
 QY 370 ValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro 389
 Db 5130 TCGGCGATCTCTTCGCGCGCGCCAGCAC-----CAGGAG 5095
 QY 390 IleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
 Db 5094 GTCCG-----ATCGCTACCGCTACCTCAAGAA----- 5065
 QY 410 SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsn----- 424
 Db 5064 ---GCGATGAACGAGCGCGCCAGCTGGCCCTGTGTCGACAAATGTGCGCGCGTGGCT 5008
 QY 425 -----LysValGlnHisTyrSerPhe 431
 Db 5007 CCGCGCTCGACCGGCATACCTACAGGACCGTACCGCGGTTACCGAGGCGCGCTTC 4948
 QY 432 PheGlyValGlnGlnAlaAsnTyrAspAsnPheThrLeuGluGlyGlyValArgValGlu 451
 Db 4947 TACATCGACACCAAGATCGATGTCGCACTGCGCACTGACCATCACCCCGCGCATCGCTCGAG 4888
 QY 452 LysGlnLysAlaSerIleArgTyrAspLysAlaLeuLe----- 464
 Db 4887 AAGATCGACAGCACTGGCGC---GACCGCGCGGCTGCTCGGCTGAAACGCGCGGTG 4831
 QY 465 -----AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArg 481
 Db 4830 CAGGAGAGAGCGCAGCAGGACTCAACGAAACCGCTGCGCGCGCTG----- 4783
 QY 482 GlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeu 501
 Db 4782 -----AGCCTGATGATCACTTTCAGCGACGAGTGG----- 4753
 QY 502 SerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGly 521

Db 4752 AAGCTGTTCGCCCACTACTCCGAATCGTTCCGCGAGCTGCAG---TACTTCCAGCTCGC 4696
 QY 522 LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer 541
 Db 4695 CAGGGCGGCGAGCGGCACTACGATACCGCGCGCGC-----CTGAGCGCGAGAGCGCC 4645
 QY 542 AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeu 561
 Db 4644 AAGACCTACGAACCTGGGCGACGCTACGACACCGAGCTGGCGCGCGAGATCAGCTG 4585
 QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
 Db 4584 TTC----- 4582
 QY 582 LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
 Db 4581 ---TACATCGACTTCGACGACGAGCTGCTAGCTCAGC---AACGACGCTCGCTGGACC 4528
 QY 602 PheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyVal 621
 Db 4527 AACCTCGCGCGC-----ACCAAGCAGCCAGGCGCATCGAGACT 4492
 QY 622 SerGlyAspTyr-----ValArgGly 628
 Db 4491 TCCGGCCACTATGACTTCGCGCTCTCGATCCGCGCTGAGCGGCTGTGCTACGCG 4432
 QY 629 ArgLeuLys-----AsnLeuProSerLeuProGlyArgGlu 640
 Db 4431 AGCTCTACCTATACCGCGCCACCTACGAAAGGCGACATTCCTCTTCAAGGGCGCGAC 4372
 QY 641 AspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPro 660
 Db 4371 CTGCGCTCTACTCGCGC-----CAGTGGCC 4345
 QY 661 AlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeu 680
 Db 4344 ACCCGCGGGTGGCTACGAGTCT-----GACCGCTGCGCATCAACCTC 4300
 QY 681 AspTyrTyrArgValPheAlaGlnAsn-----LysLeuAlaArgTyrGluThrArgThr 698
 Db 4299 GAC-----GCTTCGCGCGCTGCTGATGCGCGTTCGCTACGACTTCGCGCGCGAGGCTC 4130
 QY 699 ProGly-----HisMetLeuAsnLeuGlyAlaAsnTyrArgArg 712
 Db 4249 CCAGGCAACTTCCACCACTACACCGACCGCGCGCGCGCGCGCGCTACGCGGA 4190
 QY 713 AsnThrArg-TyrGlyGluTyrAsnTyrValLysAlaAsp----- 726
 Db 4189 CATTCGCGCTACGCTGAGTGGATGCGCGTTCGCTACGACTTCGCGCGCGAGGCTC 4130
 QY 727 -----AsnLeuLeuAsnGlnSerValTyrAlaHisSe 737
 Db 4129 GAACCTGAAGCTCGGAGTGGGTGAAGAACCTGTTGCAAGCAGTATTTCACCCCGCTC 4070
 QY 737 rSer 738
 Db 4069 CAGC 4066
 RESULT 36
 AE006093
 LOCUS
 DEFINITION Pasteurella multocida PM70 section 60 of 204 of the complete genome.
 ACCESSION AE006093
 VERSION AE006093.1 GI:12720840
 KEYWORDS Pasteurella multocida
 SOURCE Pasteurella multocida
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Pasteurella.
 REFERENCE 1 (bases 1 to 10777)
 AUTHORS May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and Kapur,V.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

Complete genomic sequence of *Pasteurella multocida*, Pm70
Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
21145866
11248100
2 (bases 1 to 10777)
Zhang, Q. and Kapur, V.
Direct submission
Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
55108, USA

FEATURES

source

Location/Qualifiers

1. .10777

/organism="Pasteurella multocida"

/mol_type="genomic DNA"

/strain="PM70"

/db_xref="taxon:747"

complement (74. .1027)

/gene="trxB"

/notes="synonym: PM0573"

complement (74. .1027)

/gene="trxB"

/codon_start=1

/transl_table=11

/product="trxB"

/protein_id="AAK02657.1"

/db_xref="GI:12720841"

/translation="MNAMKHSQILLGSGPAGVTAIYAARANKLPVLVTGLQGGQL
TTTETENPGDGETTPELMQHAERKFDTEIIPDHINKVLSRRPFRKLYGVDVH
TSCDALIATGASARYLGLPSEAYKRGVSACATCGDFYRNKPAVIGAGTAVE
EALYLANIASEVHLVRRDSFRSEKILIDRLMKVVEGKILIHTRNLTDELVDGDMGV
TGVRLKDTSEATELSLQGVFIAGHAPNTDIFAPQLELNNGYIVVKSGLGNATAT
SDGVFAAGSDVMDHNRQATISAGTCMAALDAERYLDAQ"

complement (1109. .1762)

/gene="PM0574"

complement (1109. .1762)

/gene="PM0574"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK02658.1"

/db_xref="GI:12720842"

/translation="WVAQAFRIQTLPTTVLFKEAQAALDAPPGILDEQALQORLALILP
KEDEIKFMALEWLENEFEQALPLIKAEWLSAKNSDIALIYAEVTIAMKKVSPA
DILNQIPLODRNMGLOAQIHELLQADTPEIQLOQSEPAKNSPQIALKALQLH
QANREASLLDFDLRLDNGEIKQOFLAULTAIGQNDPVSNNKRYRLLSLYL"

complement (2098. .2787)

/gene="PM0575"

complement (2098. .2787)

/gene="PM0575"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK02659.1"

/db_xref="GI:12720843"

/translation="MKKIEQFANNYSWALRMKEENSSYFKELADHQTSPYLWIGSD
SRYPACGLNLGPGELFHLNRVANQVHTDNLCLSVQAVELVIEHIIICGHTNCG
GIKAMADODLGINNWLHLDIWKVSHLLGNLPKRAMLTKINVAEOVYNLR
SSIIKSAKNGKSLSLHGWYVDYNDGFLIDQGVIASTRSLEISYRNAARLLKNE
ELIKQODEBA"

2953. .5181

/gene="hemR"

/notes="synonym: PM0576"

2953. .5181

/gene="hemR"

/codon_start=1

/transl_table=11

/product="HemR"

/protein_id="AAK02660.1"

/db_xref="GI:12720844"

/translation="MDKMLKGCVFSLVGCIGIQLASNPDPVDLPIIVNADE
DNKLPGRSVLQKXIDQQAADNADLNLPGVNMAGGFRPGSGTINMGMDAEDVR
VOLDGATKSPKQCGSIFTEPELLRKVTVDKGNYSPOYGGGAGTVPKTDATDF
LKENQKIGGLFKYGNNSNNQKTYSTALVLQNEQKNIDLLLPQSVRNASNYRPPDKSK

gene

CDS

complement (5220. .6026)

/gene="trpA"

/note="synonym: PM0577"

complement (5220. .6026)

/gene="trpA"

/codon_start=1

/transl_table=11

/product="trpA"

/protein_id="AAK02661.1"

/db_xref="GI:12720845"

/translation="MRRFDTLFSQLOQOQGAFFVFTLCPDNLASFEIIQTLVENG
ADALELGFPSDPMLDGPVIOAANORALVAGCSDASFOLLTOVRSTYPTLPISLLC
ANLYAQGLIERFYQRCHEVGDVAVLIADLPLFHREFTVMAEKYSIQQVFCPPNACE
DTIKRVAQQTQGYTILVRSRGVGVENQSQADMKLTLVERLKAYQAPPIQGGFISOP
QQVSDVLKIGVNGAIGSATVQIIEKYHDDPQACLAELAQFTQMKQAIT"

complement (6013. .7221)

/gene="trpB"

/note="synonym: PM0578"

complement (6013. .7221)

/gene="trpB"

/codon_start=1

/transl_table=11

/product="trpB"

/protein_id="AAK02662.1"

/db_xref="GI:12720846"

/translation="MSETLNPYGFEGGMVPEILMPVLKNLEKAFVEAQODPTKE
TELDLKNYAGRPALTACALACMLPGCQIYMGAKDVERQSPNVFMRMLGANVTAV
TKRSASLQACCEAMRDWENYETHYLLGTAGHPHPTTIVREFOKIIGETKQQLL
AREGLPDVAIVAGGSGNAIFNDPIETSVRLIGVEPAGGIATQGGAPLGHGT
TGIVGKMAPIQMDPGQIBESYSISAGLDFPSPQHAHLQAGRAQYBSITDDEAL
SAFOALAREHGIIPALESALAYALIKIQOPEKEQLLVNLSGRGDKDIFTVDRL
SQKSVSYAPF"

complement (7423. .8181)

/gene="ydfG"

/note="synonym: PM0579"

complement (7423. .8181)

/gene="ydfG"

/codon_start=1

/transl_table=11

/product="ydfG"

/protein_id="AAK02663.1"

/db_xref="GI:12720847"

/translation="MKKIALVTGATAGTGTATCRTLIQHNVNVIGTGRAERLAALQO
EFGADFLPLADCLDRHATEKAINNLQAAVAVLNNAGLALGDRADKALDDWI
QMTNTIGLVTTRLLPQWVRNQGHIIINLQAGTYPYGGNVYVASKAFKQFS
LNLADLQAGTAVTNTIEPGLCGSTESNVREKGDQRAAKLYVEDVQVFTPEDIANIV
LWANOQEHVNNIRIEVMPAQSSAPLNVKQTK"

complement (8217. .9626)

/gene="trpC"

/note="synonym: PM0580"

complement (8217. .9626)

/gene="trpC"

/codon_start=1

/transl_table=11

/product="trpC"

/protein_id="AAK02664.1"

/db_xref="GI:12720848"

/translation="MTQAFPSILQKIIQDKTTTAAEQALPLASFODLERSDRDPY
QAMQGSQPSYILECKKSPKSLRAAFDLTAEVYQHVATVSVLTDRKYFG
HQAIAQVRQVTPQILCKDFILISEYQVLYRAQADAILMLSVLNDEREYQALATLA
HQLGVLITETSEAFARALKARQVIGVNNRNLHLSVDIORVAQITQHYADQIFP
ETKIISSSGIYQHQIIEHLRHVADGFLIGSSLMQHTDLNHAHVRSLLFGEHKVGLTRP

gene

CDS

complement (5220. .6026)

/gene="trpA"

/note="synonym: PM0577"

complement (5220. .6026)

/gene="trpA"

/codon_start=1

/transl_table=11

/product="trpA"

/protein_id="AAK02661.1"

/db_xref="GI:12720845"

/translation="MRRFDTLFSQLOQOQGAFFVFTLCPDNLASFEIIQTLVENG
ADALELGFPSDPMLDGPVIOAANORALVAGCSDASFOLLTOVRSTYPTLPISLLC
ANLYAQGLIERFYQRCHEVGDVAVLIADLPLFHREFTVMAEKYSIQQVFCPPNACE
DTIKRVAQQTQGYTILVRSRGVGVENQSQADMKLTLVERLKAYQAPPIQGGFISOP
QQVSDVLKIGVNGAIGSATVQIIEKYHDDPQACLAELAQFTQMKQAIT"

complement (6013. .7221)

/gene="trpB"

/note="synonym: PM0578"

complement (6013. .7221)

/gene="trpB"

/codon_start=1

/transl_table=11

/product="trpB"

/protein_id="AAK02662.1"

/db_xref="GI:12720846"

/translation="MSETLNPYGFEGGMVPEILMPVLKNLEKAFVEAQODPTKE
TELDLKNYAGRPALTACALACMLPGCQIYMGAKDVERQSPNVFMRMLGANVTAV
TKRSASLQACCEAMRDWENYETHYLLGTAGHPHPTTIVREFOKIIGETKQQLL
AREGLPDVAIVAGGSGNAIFNDPIETSVRLIGVEPAGGIATQGGAPLGHGT
TGIVGKMAPIQMDPGQIBESYSISAGLDFPSPQHAHLQAGRAQYBSITDDEAL
SAFOALAREHGIIPALESALAYALIKIQOPEKEQLLVNLSGRGDKDIFTVDRL
SQKSVSYAPF"

complement (7423. .8181)

/gene="ydfG"

/note="synonym: PM0579"

complement (7423. .8181)

/gene="ydfG"

/codon_start=1

/transl_table=11

/product="ydfG"

/protein_id="AAK02663.1"

/db_xref="GI:12720847"

/translation="MKKIALVTGATAGTGTATCRTLIQHNVNVIGTGRAERLAALQO
EFGADFLPLADCLDRHATEKAINNLQAAVAVLNNAGLALGDRADKALDDWI
QMTNTIGLVTTRLLPQWVRNQGHIIINLQAGTYPYGGNVYVASKAFKQFS
LNLADLQAGTAVTNTIEPGLCGSTESNVREKGDQRAAKLYVEDVQVFTPEDIANIV
LWANOQEHVNNIRIEVMPAQSSAPLNVKQTK"

complement (8217. .9626)

/gene="trpC"

/note="synonym: PM0580"

complement (8217. .9626)

/gene="trpC"

/codon_start=1

/transl_table=11

/product="trpC"

QDVRVYQOQALYGLLI FVTHSKRCSISURQAQELVLTQAPLRFVGVFQDQSI AFICHLA
TQLRLHAVQLHGNESHEFTRQLRDELPCNCEIWRACIESSGTSPCQIVENKNITDRYI
FDSKTHQQRGTGKTDFDWQI LPHHLKPKMSLGAITPENIPLAQAQGLGLDUNSGLE
TPYGVKSHQKVVAFAOLR"

gene
IIFGVNSHQKVVAFRAQDK
complement (9663..10664)

```

gette
comp1emmc(5000
/ gene="trpD"

```

```
/gene=crp  
/note="synonym: PM0581"
```

cns
complement (9663...10664)
/note= synonym: FRO301

```
cds
complement(5000
/gene="trap"
```

```
/gene=:c1p0
/codon start=1
```

```

/cocon_state=1
/transl_table=11

```

```

/CLAUSE_CAPTION="TYPD"
/product="TYPD"

```

```

/protein id="AAK02665"
/protein id="I7PD"

```

```
/protein_id="AAK02863.1"
/db_xref="GI:12720849"
```

/wd_xref:=GI:147020645
 /translation="MTGDSLLHQLPEKKAALNQOQTEPLTAIVRGQLNSQLSALIA
 LKLVKPEETISGATVAQDAEESPDPYPADIVGTGQDANTINISTASGAAVAAS
 YGKVAKPEESVSSQTSQDALLTGVNVNHSIAHKARQALDIEGICFLPQYHUGF
 QHAIIVRQALTRITFNILGIPKAPKRVQYSSDLIOPYATVARLGHSHSV
 WSGSLDIALTHGTVTEVAIKAGQIERYTLTPQDFQFQPLETTRGKPSNAQVLT
 ALLOQKGLKHAQAVAMNSAGMSFGYENLKQNAQAIMDIATGKPFETLQOQLOY"
 0042 c 2253 a 3304 t

BASE COUNT	3178 a	2042 c	2253 g	3304 t
------------	--------	--------	--------	--------

FASE C
ORIGIN

Alignment Scores:	
Pred. No.:	1.6e-05
Score:	253.00
Percent Similarity:	35.47%
Best Local Similarity:	20.75%
Query Match:	6.27%
DB:	1
Length:	10777
Matches:	165
Conservative:	117
Mismatches:	291
Indels:	222
Gaps:	39

US-09-936-377-2 (1-758) x AE006093 (1-10777)

Qy	67	ArgGlnIysAlaValAsnLeuGlyAspAlaLeuaspGlyValProGlyIleHisAlaSer	86
Db	3133	CAACAACAAGCAGATATGCTGCTGACTATAATAATTTTGGCTGGTGTAAATATGGCG	3192
Qy	87	GlnIyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLys	106
Db	3193	-----GGAGGATTTCGCCCT-----GGTGGTCAACA-----	3219
Qy	107	ValLeuAsnHisHisGlyGluThrGlyAspMetAlaaspPheSer-----	121
Db	3220	---TTAAATATTAAATGA---ATGGTGATGCTCAAGATGTTAGACTTCAACTAGACGGC	3273
Qy	122	-----ProAspHisAlaIleMetValaspThrAlaLeuSer	133
Db	3274	GCAACAAAGAAGTTTCGAAAAATATCAACAAGGCTCTATTTTATTGAACCTGAGTATTATA	3333
Qy	134	GlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAla	153
Db	3334	AGAAAGGTGACAGTAGACAAAGGAAATATTCTCTCAATATGCCAATGGTGGCTTTGCT	3393
Qy	154	GlyLeuValaspValAlaaspGlyLysIleProGluLysMetProGluAsn---GlyVal	172
Db	3394	GGTACTGTAAATTTTGAAACAAAAGATGCAACTGATTTTTTGGAAAGAAATTCAGAAAAATA	3453
Qy	173	SerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGly	192
Db	3454	GGTGGATTATTAAATATGGAATAATAGCAATATAACCAAGAAACCTTATAGTACAGCC	3513
Qy	193	IleAsnIle---GlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys	211
Db	3514	CTAGTTTTACAGATGAACAATAATATTCATTTGTTATTATTATTGGTCTGTGAAGAAT	3573
Qy	212	SerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgProAspSerProArg	231
Db	3574	GCAAGCAATTATACAGACCT-----GATAAAGTAAAAATCTTTTTTCAAAAAC	3624
Qy	232	ArgPheAlaasnGly-----GlnHisArgAlaVal	241
Db	3625	AATCAAAAAGGTGATTAATAAAGTAAATTTGGCAAAATTACTCTGCAACATTATTAACT	3684

Db	FEATURES	Location/Qualifiers
4552	CTATTCAGAAATGATGCTTTTCATATTAGAACTACTTATTATTTCAATCGCGCAAG 4611	1..10399 /organism="Plesiomonas shigelloides" /mol_type="genomic DNA" /db_xref="taxon:703" complement (386..943) /gene="hug2" complement (386..943) /gene="hug2" /note="heme utilization protein"
568	AsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProlyssSerIleGluAspAsp 587	/codon_start=1 /transl_table=11 /product="Hug2" /protein_id="AAG23392.1" /db_xref="GI:13774059"
4612	AATGAATCTTCAAAACG-----AGAGGGTTAATTTCGCTAGAAAATGCT 4656	/translation="MKSDIVMTNERQELQNLQPEIAEPNECRTLVLATVDADGMP RVSTAPVQLTDGTYVLISEIARHARNLQVPTSLMMIEDSESQRLFAKRLTFDA NVEKVARDTASQBAVALEARHGDIVANLSQMAADFILFRMVPEQGLYVKGQAFQV SSDLVDVHLQEGHKTSGEPVKA" complement (1108..1590) /gene="hugX" complement (1108..1590) /gene="hugX"
588	SerGluMet-----LysLeuValArg-----TyrAsnGlnSerGly 599	/note="heme utilization protein"
4657	TTAGATGTTGATATAGTGTGTCCTAAATATTAGTAATTACCGTAATTACTCGT 4716	/codon_start=1 /transl_table=11 /product="Hug2" /protein_id="AAG23392.1" /db_xref="GI:13774059"
600	AlaAspPheTyrAlaGluGlyGlyLeuTyrPheLysProThrArgTyrArgIle 619	/translation="MKSDIVMTNERQELQNLQPEIAEPNECRTLVLATVDADGMP RVSTAPVQLTDGTYVLISEIARHARNLQVPTSLMMIEDSESQRLFAKRLTFDA NVEKVARDTASQBAVALEARHGDIVANLSQMAADFILFRMVPEQGLYVKGQAFQV SSDLVDVHLQEGHKTSGEPVKA" complement (1108..1590) /gene="hugX" complement (1108..1590) /gene="hugX"
4717	TATGTTATCAAGGAGCGAATTAGAAGCTTATTATCAATCATCGTATTATTGCT 4773	/note="heme utilization protein"
620	GlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArg 639	/codon_start=1 /transl_table=11 /product="HugX" /protein_id="AAG23393.1" /db_xref="GI:13774060"
4774	GGACTGCATATCTTATGTAAAGAAACGGCATCTTCCACAAAGAAATCCATGGAGT 4833	/translation="MSASTQBLAARFGVSEYQVLAALPEAMVMAHGSWAQSLAEELT DWGPUTTIVSESGIFEPFGAPFVKGPGHFNLYSREGTGLHGLHLLSQVSHIALLA KPFMGKASYSLOFFAANGRTMFKYLGKNGKORLLPEQVLFQELRAKLIQAQPPAAE " complement (1814..3169) /gene="hugW" complement (1814..3169) /gene="hugW"
640	GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgVal 659	/note="heme utilization protein"
4834	AAACATCT-----ACATGATCGCA-----GAAACACCATAGAAA 4872	/codon_start=1 /transl_table=11 /product="HugX" /protein_id="AAG23393.1" /db_xref="GI:13774060"
660	ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679	/translation="MSASTQBLAARFGVSEYQVLAALPEAMVMAHGSWAQSLAEELT DWGPUTTIVSESGIFEPFGAPFVKGPGHFNLYSREGTGLHGLHLLSQVSHIALLA KPFMGKASYSLOFFAANGRTMFKYLGKNGKORLLPEQVLFQELRAKLIQAQPPAAE " complement (1814..3169) /gene="hugW" complement (1814..3169) /gene="hugW"
4873	GCAACCGCTACTTTAGGTTTAAATATTCGGAATAT----- 4908	/note="heme utilization protein"
680	LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699	/codon_start=1 /transl_table=11 /product="HugW" /protein_id="AAG23394.1" /db_xref="GI:13774061"
4909	-----TATTTACGGCAGGTGCGTGTGAGTTTGTGAGAACGATAGATACCA 4962	/translation="MMIDLSVLDSDILGHSPDPLAVALFTGKQAHAGSGRPLPPP AVGQQLQMLQMDTPAIRPCLYIHIFPCRVCTPCSFQYAAASLTVAETFAALUSEE LKRKAAMPWTQAAPFAAYIIGGTPDTLTAEQIFALQGLIKAHPELTDCCITBEGRL NRFSEKTHAAVDGVRNFRSFGVQSDTDVRRKARLDPREKLETLVSHVLRAPAI IVDLMFGLPHQSATWQDLQDILHSGAHGVDLYOLIEMOQTPMARLVEENLRQAPAA TPEKSRMVAQGVAFMCOHQIQQISVNHVTRDTRERSYVNHAKTTADVLPAAGAGGF IDGFALMOTRDLSEYHOAIRSERVPMAMSPNPOHALHCAIKAGFDRGLQRSELD LSGMALYIVSPFLTQWQERGLLSLEADQARLTAGSFNINMAQGLINLRLESPPLA A" complement (3169..5235) /gene="hugA" complement (3169..5235) /gene="hugA"
700	-----GlyHisMetLeuAsnLeuGlyAlaAsnTyrArgArg 712	/note="heme receptor"
4963	CTATCTAATGATCTAAAGCATCATATTGGGCATTACCTTCTTCAAGGATATAGCCTA 5022	/codon_start=1 /transl_table=11 /product="HugW" /protein_id="AAG23395.1" /db_xref="GI:13774062"
713	AsnThrArgTyrGlyGluTrp-----AsnTyrTyrValLys 724	/translation="MPVQPSAIFRSTVSATAAMTAKHTGFRKSALSLLILNSVAVSF ANAAAPSPOTQNLNEVLATNTRENKALSQETRSVAVVTAQOIQQOQPSVAEALKYIP NVDVSGGFRANQSPVSGLSNRILQVVDGVRQNTTSGHRLNSWDEMLSSIEVIX GPTSLWGSGLGVAQQTISAPDMLHGDQSVGGYLRQGVASANDETKTSQAVTVGLL GDSVDFLLNGYADANNRLNGEHLTDSAYRNOGLAKFGWQADQRLSLVRQAL TRQNAFSPQSVMTSVPLVQOKTRDFNTLDYRLNPTDSRWLDARVAYVNTKTEFDE YRVTKQSDKVDYRLTGLNINSHSQFDLLSLTYGGDIYEDKTOGRECKNRPIDAGR SKVMSGYQADIPLGSONLLPGLYDHFTAEKNIAGSERSEHLSPSVGLRWAQR WLTINARDEAFAPSMEMETTTGTHFCMGPGCNVFFANFDLPETAKKEISANQR FNWVLADLAFGATYFNNVSNYIDQQNMFTTQYVNVTDARLNGVLDKARNWRL ETSLSYAQTEGDKTKNQPLNNIPAKRWLVGVSFHYFMDRLKAGVNVSHYSQDELPS
5023	CATAGTGTATCTCTCTGAGTCTCAAGAGATTAAAGAAATGAAATTTCAAGTTACT 5082	/codon_start=1 /transl_table=11 /product="HugA" /protein_id="AAG23395.1" /db_xref="GI:13774062"
725	AlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerPheLeuSerAspThrPro 744	/translation="MPVQPSAIFRSTVSATAAMTAKHTGFRKSALSLLILNSVAVSF ANAAAPSPOTQNLNEVLATNTRENKALSQETRSVAVVTAQOIQQOQPSVAEALKYIP NVDVSGGFRANQSPVSGLSNRILQVVDGVRQNTTSGHRLNSWDEMLSSIEVIX GPTSLWGSGLGVAQQTISAPDMLHGDQSVGGYLRQGVASANDETKTSQAVTVGLL GDSVDFLLNGYADANNRLNGEHLTDSAYRNOGLAKFGWQADQRLSLVRQAL TRQNAFSPQSVMTSVPLVQOKTRDFNTLDYRLNPTDSRWLDARVAYVNTKTEFDE YRVTKQSDKVDYRLTGLNINSHSQFDLLSLTYGGDIYEDKTOGRECKNRPIDAGR SKVMSGYQADIPLGSONLLPGLYDHFTAEKNIAGSERSEHLSPSVGLRWAQR WLTINARDEAFAPSMEMETTTGTHFCMGPGCNVFFANFDLPETAKKEISANQR FNWVLADLAFGATYFNNVSNYIDQQNMFTTQYVNVTDARLNGVLDKARNWRL ETSLSYAQTEGDKTKNQPLNNIPAKRWLVGVSFHYFMDRLKAGVNVSHYSQDELPS
5083	GTTGATATATTTATTAACCGACCTATAT-----CCT 5115	/note="heme utilization protein"
745	GlnMetGlyArgSerPheThrGly---GlyValAsnValLysPhe 758	/codon_start=1 /transl_table=11 /product="HugA" /protein_id="AAG23395.1" /db_xref="GI:13774062"
5116	TACTTAGGAGAAATAGCTTCAGAACAGTAGGAGTATGTCAAATTT 5160	/translation="MPVQPSAIFRSTVSATAAMTAKHTGFRKSALSLLILNSVAVSF ANAAAPSPOTQNLNEVLATNTRENKALSQETRSVAVVTAQOIQQOQPSVAEALKYIP NVDVSGGFRANQSPVSGLSNRILQVVDGVRQNTTSGHRLNSWDEMLSSIEVIX GPTSLWGSGLGVAQQTISAPDMLHGDQSVGGYLRQGVASANDETKTSQAVTVGLL GDSVDFLLNGYADANNRLNGEHLTDSAYRNOGLAKFGWQADQRLSLVRQAL TRQNAFSPQSVMTSVPLVQOKTRDFNTLDYRLNPTDSRWLDARVAYVNTKTEFDE YRVTKQSDKVDYRLTGLNINSHSQFDLLSLTYGGDIYEDKTOGRECKNRPIDAGR SKVMSGYQADIPLGSONLLPGLYDHFTAEKNIAGSERSEHLSPSVGLRWAQR WLTINARDEAFAPSMEMETTTGTHFCMGPGCNVFFANFDLPETAKKEISANQR FNWVLADLAFGATYFNNVSNYIDQQNMFTTQYVNVTDARLNGVLDKARNWRL ETSLSYAQTEGDKTKNQPLNNIPAKRWLVGVSFHYFMDRLKAGVNVSHYSQDELPS
RESULT 37		
AY008342/c		
LOCUS	10399 bp DNA linear BCT 23-APR-2001	
DEFINITION	Plesiomonas shigelloides heme iron utilization locus, complete sequence.	
ACCESSION	AY008342	
VERSION	AY008342.1 GI:13774058	
KEYWORDS	Plesiomonas shigelloides	
SOURCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
ORGANISM	Enterobacteriaceae; Plesiomonas.	
REFERENCE	1 (bases 1 to 10399)	
AUTHORS	Henderson,D.P., Wyckoff,E.B., Rashidi,C.E., Verlei,H. and Oldham,A.I.	
TITLE	Characterization of the Plesiomonas shigelloides genes encoding the heme iron utilization system	
JOURNAL	J. Bacteriol. 183 (9), 2715-2723 (2001)	
MEDLINE	21189235	
PubMed	11292789	
REFERENCE	2 (bases 1 to 10399)	
AUTHORS	Henderson,D.P., Wyckoff,E.B. and Payne,S.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (29-SEP-2000) Science and Mathematics, University of Texas of the Permian Basin, E. University Blvd, Odessa, TX 79762, USA	


```

QY 183 GlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnVal 202
Db 4626 GCCAATGATGAACCAACCAACGCGGTGCGTATGCGCTGTTGGGGGATTCGCGCAT 4567
QY 203 LeuHisThrGluGlyLeuThrArgLysSerGlyAspTyrAlaValProArgTyrArgAsn 222
Db 4566 TTCCTGCTAACCGCTACTACGCGATGCC-----ATAAC 4531
QY 223 LeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeu 242
Db 4530 CTG-----CGTTGGCAATGGCGAGCAC----- 4507
QY 243 GlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeu 262
Db 4506 -----CTGACTGACTCGGCCTATCGCAATCAGGCGGGTTA 4471
QY 263 ProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeu 282
Db 4470 GCCAAATTTGGCTGGCAGCGAGATGACGCG-----CAGCGTCTGGAG 4429
QY 283 IleAsnLysArgTyrLeuGlnLeuTyrProHis----- 293
Db 4428 CTGTCAGTTCCGCAACCGCAACCCATCAGAACGCCAGCAGCAGCAACCCATCAGATGGTC 4369
QY 294 -----LeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSer 308
Db 4368 AGCACCTCAGTGGCGCTGGTGCAGCAGAAACCGTGATTTCC--AATACCACCTTGGAT 4312
QY 309 CysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTrpIle 328
Db 4311 TATCGCTCATCCCAACCGATAGCCGT-----TGGTTG 4279
QY 329 AspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPhe 348
Db 4278 GACCGACGCGTAACCTGCTATGTGTGACAAACAAACGAA-----TTT 4240
QY 349 GluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAsp 368
Db 4239 GATGAATATCGCTCAGCAAAAGCAAGCAAGGATAAGGTGATTACCGCACACCTGGGTCTG 4180
QY 369 AlaValGluAsn-----PhePhe 374
Db 4179 AATATCAGTAAACCAATCGCAGTTGATCTGCTGAGCCTGACCTATGCGCGCATATTTAT 4120
QY 375 AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle-----Gly 391
Db 4119 GAAGATAAAACCCAAAGCGAGCGG---GAAGGGAAGAAACCGCGCGATTCACCGCATGGT 4063
QY 392 ArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerAlaLeuSerAla 411
Db 4062 GCGAGCAAA---GTCTGGGA-----AGTTATGACAGGCC 4030
QY 412 ThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhe 431
Db 4029 GATATTCCATTAGGTAGTCACTGAGCAACTGTTACCGGCT----- 3991
QY 432 PheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGlu 451
Db 3990 -----CTGCGTTATGACCCCTTTACCGCGAA----- 3964
QY 452 LysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys 471
Db 3963 -----GACAAAGAATATCGCGGCTCCGAGCGTAGCGAGAT 3928
QY 472 GlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGly 491
Db 3927 CACCTGTGCGCATCGTTGGCGCT----- 3904
QY 492 AsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu----- 509
Db 3903 CGTTGGCGGCGGACTGATGG-----CTGACCTTAATGCTCCCTTACGAGCAAGCGCTTC 3850

```

```

QY 510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis-----Val 524
Db 3849 CGTCACCGAGCATGGAGAGATGTACACCAACCGCTACTCTCTCGATGGGCCCGCATG 3790
QY 525 AlaThrAsnThrPheGluValGlyLeuAsnLysHisLeuAsnLysGluArgSerAsnAsnIle 544
Db 3789 GGCTGCAACGTTTAA---GCCAATCCCGATTGAACCGGAGACCGCCAAACAAA 3733
QY 545 GluLeuAlaLeuGlyTyrGlu-----GlyAspArgTrpGlnTyrAsn 558
Db 3732 GAGATCTTCAGCTCAGATCGCGCTTTAACAAACGTCGTCAGATGATGATAGCGTTTGGG 3673
QY 559 LeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGln----- 573
Db 3672 GCGACTTACTCCATAACACGTCAGCACTACATTGACACGAGCAGCAACATGTTCCACC 3613
QY 574 -----ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGlu 589
Db 3612 ACCCAATACGTGAACGTCAGCGATCGCGT---TTGCGGGCGCTCGAGCTGATGCG--- 3559
QY 590 MetLysLeuValArgTyrAsn-----GlnSerGlyAlaAspPheTyrGlyAla 605
Db 3558 -----CGTTTACTGCGCGCATCTGGAACACGACCTGAGCTATGCCAGACA 3511
QY 606 GluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyr 625
Db 3510 GAAAGG----- 3505
QY 626 ValArgGlyArgLeuLysAsnLeuPro-----SerLeuProGlyArgGluAspAlaTyr 643
Db 3504 ---CGCGACAGAAACCAACGCGCTGATACATTCGCGCATTAAGTGGGTATTG 3448
QY 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 3447 GCGCGTGGCCACTACTTATGATCGCGAC----- 3418
QY 664 LeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyr 683
Db 3417 -----CTGAAAGCGGCGGT-----AACGTGACGCCATTAC 3388
QY 684 ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArg-----ThrPr 699
Db 3387 GAA-----TCACAAGATGAATGCCATCTAATGTGACCAATGTATCCGCTCTACACC 3334
QY 699 OGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyr 719
Db 3333 TTGGTGGATCTCTACACCACTTGGCAGCCACAAACGCGCGCTCAAAGCAGTCAAAGTG 3274
QY 719 pAsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPh 739
Db 3273 GATGGGT-ATC-----GATACGTGACTGACGAA-----TATTACCGTCAGCGCTT 3227
QY 739 eLeuSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 3226 T---GACCAGCTCTATTCCGCGGTGCTAACTTAAGTTGGTGGTGGTATTCCTTC 3172

```

RESULT 38

AE015838

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AE015838 12558 bp DNA linear BCT 02-DEC-2002
 Shewanella oneidensis MR-1 section 387 of 457 of the complete genome.

AE015838.1 GI:24350433

Shewanella oneidensis MR-1

Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

1 (bases 1 to 12558)

Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, W.,

Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R.,

Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S.,

DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R.,

Peterson, J., Umayam, L., White, O., Wolf, A., Wolf, A., Vamathevan, J., Weidman, J., Imprim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.
Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*
Nat. Biotechnol. 20 (11), 1118-1123 (2002)
22297686
12368813
2 (bases 1 to 12558)
Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, W., Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R., Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R., Peterson, J., Umayam, L., White, O., Wolf, A., Vamathevan, J., Weidman, J., Imprim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.
Direct Submission
Submitted (12-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
FEATURES
source
1..12558
/organism="Shewanella oneidensis MR-1"
/mol_type="genomic DNA"
/strain="MR-1"
/db_xref="taxon:2111586"
complement(183..335)
/locus_tag="SO4065"
complement(183..335)
/locus_tag="SO4065"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057039.1"
/db_xref="GI:24350434"
/translation="MCLNVRCLATFLLTNSLLSLRKALKVKRANKSGFSGVDNKGGRDSG"
389..1492
/locus_tag="SO4066"
389..1492
/locus_tag="SO4066"
/note="identified by match to PFAM protein family HMM PF01259"
/codon_start=1
/transl_table=11
/product="phosphoribosylaminoimidazole-succinocarboxamide synthase, putative"
/protein_id="AA057040.1"
/db_xref="GI:24350435"
/translation="MSLADSVLAVNNDLPRTDKSVHSGKVRVYVWLTADSRRLIKT KGNVPDETLAVISDRISAPDCIFHGGGLKIPCKGAALNAISNHWFKLFAENG LADSHLDIPHPFWVQKARPKEVAICQYITGSMWRAYSGERVFCGIIPEGLE KDQPELLITPTKGLITGIPGVAODDNIERSIDIANQAPGEKVEDIDLXEKL KDQFVTSKALADLDQVVDVTFEFGYVTDQDNGSKLIYMDVGTDPDSRWDGAA RDGKILNKGSRFQFLNHPDPDILLNDRMPERBALDARNALPLEAMMQVSRITYT GIAEKTGAALPLPANKPAKIIKILREYDLIV"
complement(1828..1974)
/locus_tag="SO4067"
complement(1828..1974)
/locus_tag="SO4067"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057041.1"
/db_xref="GI:24350436"
/translation="MLNFKLVFGITLFAHYHQATPKPKQYTFPCPEVAGRWKRLRE SGHG"
2140..4248
/locus_tag="SO4068"
2140..4248

Peterson, J., Umayam, L., White, O., Wolf, A., Wolf, A., Vamathevan, J., Weidman, J., Imprim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.
Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*
Nat. Biotechnol. 20 (11), 1118-1123 (2002)
22297686
12368813
2 (bases 1 to 12558)
Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, W., Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R., Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R., Peterson, J., Umayam, L., White, O., Wolf, A., Vamathevan, J., Weidman, J., Imprim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.
Direct Submission
Submitted (12-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
FEATURES
source
1..12558
/organism="Shewanella oneidensis MR-1"
/mol_type="genomic DNA"
/strain="MR-1"
/db_xref="taxon:2111586"
complement(183..335)
/locus_tag="SO4065"
complement(183..335)
/locus_tag="SO4065"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057039.1"
/db_xref="GI:24350434"
/translation="MCLNVRCLATFLLTNSLLSLRKALKVKRANKSGFSGVDNKGGRDSG"
389..1492
/locus_tag="SO4066"
389..1492
/locus_tag="SO4066"
/note="identified by match to PFAM protein family HMM PF01259"
/codon_start=1
/transl_table=11
/product="phosphoribosylaminoimidazole-succinocarboxamide synthase, putative"
/protein_id="AA057040.1"
/db_xref="GI:24350435"
/translation="MSLADSVLAVNNDLPRTDKSVHSGKVRVYVWLTADSRRLIKT KGNVPDETLAVISDRISAPDCIFHGGGLKIPCKGAALNAISNHWFKLFAENG LADSHLDIPHPFWVQKARPKEVAICQYITGSMWRAYSGERVFCGIIPEGLE KDQPELLITPTKGLITGIPGVAODDNIERSIDIANQAPGEKVEDIDLXEKL KDQFVTSKALADLDQVVDVTFEFGYVTDQDNGSKLIYMDVGTDPDSRWDGAA RDGKILNKGSRFQFLNHPDPDILLNDRMPERBALDARNALPLEAMMQVSRITYT GIAEKTGAALPLPANKPAKIIKILREYDLIV"
complement(1828..1974)
/locus_tag="SO4067"
complement(1828..1974)
/locus_tag="SO4067"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057041.1"
/db_xref="GI:24350436"
/translation="MLNFKLVFGITLFAHYHQATPKPKQYTFPCPEVAGRWKRLRE SGHG"
2140..4248
/locus_tag="SO4068"
2140..4248

Peterson, J., Umayam, L., White, O., Wolf, A., Wolf, A., Vamathevan, J., Weidman, J., Imprim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.
Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*
Nat. Biotechnol. 20 (11), 1118-1123 (2002)
22297686
12368813
2 (bases 1 to 12558)
Heidelberg, J., Paulsen, I., Nelson, K., Gaidos, E., Nelson, W., Read, T., Eisen, J., Seshadri, R., Ward, N., Methe, B., Clayton, R., Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R., Dodson, R., Durkin, A., Haft, D., Kolonay, J., Madupu, R., Peterson, J., Umayam, L., White, O., Wolf, A., Vamathevan, J., Weidman, J., Imprim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T., McDonald, L., Feldblyum, T., Smith, H., Venter, J., Nealon, K. and Fraser, C.
Direct Submission
Submitted (12-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
FEATURES
source
1..12558
/organism="Shewanella oneidensis MR-1"
/mol_type="genomic DNA"
/strain="MR-1"
/db_xref="taxon:2111586"
complement(183..335)
/locus_tag="SO4065"
complement(183..335)
/locus_tag="SO4065"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057039.1"
/db_xref="GI:24350434"
/translation="MCLNVRCLATFLLTNSLLSLRKALKVKRANKSGFSGVDNKGGRDSG"
389..1492
/locus_tag="SO4066"
389..1492
/locus_tag="SO4066"
/note="identified by match to PFAM protein family HMM PF01259"
/codon_start=1
/transl_table=11
/product="phosphoribosylaminoimidazole-succinocarboxamide synthase, putative"
/protein_id="AA057040.1"
/db_xref="GI:24350435"
/translation="MSLADSVLAVNNDLPRTDKSVHSGKVRVYVWLTADSRRLIKT KGNVPDETLAVISDRISAPDCIFHGGGLKIPCKGAALNAISNHWFKLFAENG LADSHLDIPHPFWVQKARPKEVAICQYITGSMWRAYSGERVFCGIIPEGLE KDQPELLITPTKGLITGIPGVAODDNIERSIDIANQAPGEKVEDIDLXEKL KDQFVTSKALADLDQVVDVTFEFGYVTDQDNGSKLIYMDVGTDPDSRWDGAA RDGKILNKGSRFQFLNHPDPDILLNDRMPERBALDARNALPLEAMMQVSRITYT GIAEKTGAALPLPANKPAKIIKILREYDLIV"
complement(1828..1974)
/locus_tag="SO4067"
complement(1828..1974)
/locus_tag="SO4067"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057041.1"
/db_xref="GI:24350436"
/translation="MLNFKLVFGITLFAHYHQATPKPKQYTFPCPEVAGRWKRLRE SGHG"
2140..4248
/locus_tag="SO4068"
2140..4248

/locus_tag="SO4068"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057042.1"
/db_xref="GI:24350437"
/translation="NNKRDIEHLLSLAVSFQSYQNTKNTDIIIDVAAFIETATIKRALK SFTCPBLTAHVNSEFTPEGVISTSLAKISHLEATPKHTYIYTPDQOESNSK MLLEISNAEKAKALLVDLEGYISSKIEKSESEVDLIHSFTFIQDNTANGKFSF HISAFTIENSEKADFEHQVEIKDGSIISESEIRFNINISSETGSENEITILLGTILIF DLKALPHVAMVLTNGCESRFDLLKKKELIRLEVKIKPLEDSFFSEENKYNII NSGIIINEIOSDPNLMENHINELYNILIRKGDSSRFDRITHIIVNTKYNIAWS SSIKKEGVEPLATNFMISKFWKLNRFQGGGITSFDVITKAKILVSSKNSKVS KRYTEQKVSNTLDKETATILAECSRDTKPEIKAVETOMLSLIEDSSIDAYL RKNKLEQINSRESNAKMKLVHERENDINIKTLLESKLTTKLYQEKQSTTQ QLTFFESKRTDINKTIRSKTKLITLITDIIVPSIYLLSLEDIEKTLNLIKISGA TTISTFIWFGDMIKILNRTISKFDNIHLNKKFEKKGLTDEIILKQUNDELSKLDN DISNIDQDIKILELEEQN"
complement(4733..5518)
/locus_tag="SO4069"
complement(4733..5518)
/locus_tag="SO4069"
/note="identified by match to PFAM protein family HMM PF01209"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA057043.1"
/db_xref="GI:24350438"
/translation="MSSNALYTDLSGYDLMCSDDINYOQOASIHRLHOFGLNNGKH LDIACGPHVRHFIDLGQCSGLDINQPMLDNMRCPAQNLQDMTAFYLEQPYD LITCFYLSIHSAGIEKLSKISAVHRALMDGGIFCFNSVDKHCIDNRLUSIKHFAFE YHAFAPSGWHYNGQKQSLNLSIAKQQPFTQLTDFEADTSLHQTLMQDQHTWA TSFAELQALLAPYFEVHIFEHYDKIYVPDLSGNGALFVCVKI"
complement(5783..6496)
/locus_tag="SO4070"
complement(5783..6496)
/locus_tag="SO4070"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA057044.1"
/db_xref="GI:24350439"
/translation="MKHSLLAVALVAGGLFSGAFTATTTLAAEVDPPIHETWGVQIQVE PDMAELNVEAVTEKDAKADNSDAAVAKFIARLAAAGSKHIGIQAANLQFQYII AQDKAPBLTGTIASRQITVTITSLRLNTIVDSIAIEGMNRVNNIALSKSEAYYAK ARQBAIVDAQKAKSLAGFGENLKIWIQIYRDQRPVQVPMLEMMNAKADYAMAAES YQYGVQVITEDRVEVYVKL"
complement(6671..6796)
/locus_tag="SO4071"
complement(6671..6796)
/locus_tag="SO4071"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA057045.1"
/db_xref="GI:24350440"
/translation="WISQDGVSTSYSPHNAVFRYQSAFRPAGIHFRARSVQGR"
6888..8333
/gene="y11g"
/locus_tag="SO4072"
6888..8333
/gene="y11g"
/locus_tag="SO4072"
/note="identified by match to TIGR protein family HMM TIGR01579"
/codon_start=1
/transl_table=11
/product="MiaB-like putative RNA modifying enzyme y11g"

protein id="AAN57046.1"
 /db_xref="GI:24350441"
 /translations="MWVEPHKQTTTLETPAKTLBAASADSVNTGNVATGNRIGFVS
 LGCKNLVDSEIRLTOLRIDGVNTNSYDNLVIVNTCGFDAAVEESLDVRAEAL
 ENGVLTGCLGAKENQIREVHDEITGPHSYEAVLKHVKYVPKPHNPFTSLIP
 OTGKLTFRHYAVLKISEGCDNCTECIIPALRGDLDSRGASVLDKARLVEAGVOE
 ILVVSQTSAYGKDKGRIDFNWGMFVKQDITSLARQLGMGANVRLHVIYYPWVDO
 LIPMAEGLLPLVDIPMOHASPRILKMKRPGKRVQLEAFQRMREICPDVIRKSTP
 IVGPGTEDEFEMLLDFLEEARLDRVGCFCYSEVEGAVANTIAELISEVKEEDRYR
 FMEVQASISERLARFVGRMTDILIDVDDEGAIGRSFADAPEIDGMVFINGETLEP
 GMLRVATHSDEHDLWAEVLVDADADEVEA"
 complement (8796..9044)
 /locus_tag="S04073"
 complement (8796..9044)
 /locus_tag="S04073"
 /note="identified by match to PFAM protein family HMM
 PF03889"

gene

CDS
 /product="conserved hypothetical protein"
 /protein_id="AAN57047.1"
 /db_xref="GI:24350442"
 /translations="MTKVNVAKVAEGLVSHENGRTITDNLKALVSLPFMRTEK
 PKKGYSNRKGRKAFKNSREGYQHKGDTPFGMGWVF"
 9290..9385
 /locus_tag="S04074"
 9290..9385
 /locus_tag="S04074"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAN57048.1"
 /db_xref="GI:24350443"
 /translations="MWITGEGDHFVAVCSYLSAVTESFLPIRYPIP"
 9426..10163
 /locus_tag="S04075"
 9426..10163
 /locus_tag="S04075"

gene

CDS
 Alignment Scores:
 Pred. No.: 6.36e-05 Length: 12558
 Score: 244.50 Matches: 180
 Percent Similarity: 34.7% Conservative: 114
 Best Local Similarity: 21.28% Mismatches: 317
 Query Match: 6.06% Indels: 236
 DB: 1 Gaps: 35
 US-09-936-377-2 (1-758) x AE015838 (1-12558)

QY 6 LeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu----- 20
 Db 10363 ATATCCCATGCTTTTAAATGCTCGGAA-ATATCCAGCGACTTACCTTGCAATTAGGC 10421
 QY 21 -----LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThr 35
 Db 10422 TTATCTTTGCCCCGCTGGCAATGGCAATGATAATGAACAGACT---CAGATTGAGCGT 10478
 QY 36 ValThrValValGlySerArgProArgAlaThrSerGlyLeuLeuHisThrSerThr 55
 Db 10479 ATTTTGGTCCAGGTGAG----- 10496
 QY 56 AlaSerAspIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAsp 75
 Db 10497 -----CAGAGTATCTCGCGCAATGCCCTAGGCTCAGCCGAT 10532
 QY 76 AlaLeuAspGlyValProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAlaPro 95
 Db 10533 GCCTTACTTAAAGATCAAGCGCTGATTCTCTGAAGCTGGAGAGTTCGCGCGCTACCA 10592
 QY 96 ValIleArgGlyGlnThrGlyArgGlnIleLysValLeuAsnHisHisGlyGluThrGly 115
 Db 10593 GTGCTGAACGGCATGATGGCGATCGCATTAAAGTGTTA-----GTC 10634

QY 116 AspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSer----- 133
 Db 10635 GACGGCGCAGACATTACCGCTCTCGCCCAACAAATGAACCCACCGCTATCTTATGTT 10694
 QY 134 -----GlnGlnValIleLeuArgGlyProValThrLeuLeuTyrSer 148
 Db 10695 TCTGCCAACCAAACTCACTCGTCTGAAGTGGTCGGCGCGGTGTCCTCCCGTCAGTGGCGGC 10754
 QY 149 SerGlyAsnValAlaGlyLeuValAspVal-----AlaAspGlyLysIleProGluLys 166
 Db 10755 GCGCACATATTGCTGGCGTGATTAAGGTTAATAGTCTCAATCTTAAGTATTACTGACAGC 10814
 QY 167 MetProGluAsnGlyValSerGlyGluLeu-----GlyLeuArgLeuSerSerGlyAsn 184
 Db 10815 GAAACTTAAGTTTGAAGCGGTGAGATTTCATCGGGATACCGCAGCAGCAGCAGCAGC 10874
 QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
 Db 10875 TTATTACTCGCGCCCAAGCCGGTATT-----GCCAGCAAAATGTAAAGCTTAAGC 10925
 QY 205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys 224
 Db 10926 TATCAAGGAGCCTATGAAGACGCCAACAGTTAC-----CAGCATGCTAACCGGCAC 10976
 QY 225 ArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyr 244
 Db 10977 AAAGTGCTGATACCTTTATATCGCGCTCAAAAC-----CACGCTTAACGCCAGCATGG 11030
 QY 245 ArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGln----- 259
 Db 11031 CGC-----GATGAAAACCAACCACTGGCGGTTHAACTG 11063
 QY 260 -----TyrGlyLeuProAlaHisSerHisGluTyrAspAspCys 272
 Db 11064 ACCCATCAAGCCATTCATTTCAAGGTTTCTCAACCAATACATGATGATGATGATGAT 11120
 QY 273 HisAlaAspIleIleTyrGlnLysSerLeuLeuAsnLysArgTyrGlyLeuGlnLeuTyrPro 292
 Db 11121 -----AACAAAGACTACGGTGGCTTGGTTCGC 11147
 QY 293 HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHis 312
 Db 11148 TATCTACGG-----GATTTAGAATATGACGGC-----GAGTTTAGT 11183
 QY 313 AspAspAspAlaHisAlaHisAlaHis-----Asn 323
 Db 11184 GCCAGCTCAACTGGCAGCAGTGTAAACATGAATGGGCTTCTCACCCCAAGAAAAACC 11243
 QY 324 GlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGln 343
 Db 11244 GGCATAATGCCGATGAATACAGAGGTAGCGATTACAGTACCACTGACCTGGCGCTTA 11303
 QY 344 ProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAsp 363
 Db 11304 ACTATGGGGGATGACAGTACGCTACTGCTTGGCCAGGAATACTATAGCTATCAATTAGT 11363
 QY 364 Glu-----LysAlaGlyAspAlaValGluAsnPhePhe 374
 Db 11364 GATATTGGCCAGCGGTGCGGACTATGATGGCCCACTATCAATATCAATATTAAAC 11423
 QY 375 AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLys 394
 Db 11424 GATGGTGAACGTCGCGCGCTGCAGTT-----TAC 11453
 QY 395 GlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGlu 414
 Db 11454 GGTGAATGG-----CAACAAAACCTCAATCCCGCTTGGGGTGTATCCGCT 11498
 QY 415 AlaValGlyGlnProMetLeuLeuAspAsn-----LysValGlnHisTyrSerPhePhe 432
 Db 11499 GGGGTACGCTGATGAATATGTGACCCAGATACGGGTGAAGTGCAGGCTTACAGC----- 11552

Qy 433 GlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGly-----GlyValArgVal 450
 Db 11553 -----AAATGCAATGATGGCATGCCCAATGTGGATGCC 11588
 Qy 451 GluLysGlnLysAlaSerIleArgTyrAspLysAla-----LeuIleAspArg 466
 Db 11589 GAGCGGCAAAAGCCTTTAATGCCATGCAGACAGACGACGATACCTTAATCGATGCG 11648
 Qy 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSer 486
 Db 11648 ----- 11648
 Qy 487 PheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSer 506
 Db 11649 ---ACATTGCTGGCAGCGTCAATATGTCGGCCAAACAACTGGAATTTGGCTTAGCC 11705
 Qy 507 HisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThr 526
 Db 11706 CGTAAATAATCGGCCCTTAACCTTATGAGCGTTACAGTTGGGGCCCGCGCTCATGGCA 11765
 Qy 527 AsnThrPhe-----GluValGlyAsnLysHisLeuAsn 537
 Db 11766 ACCACTATGATGGCTGATGGCAGCGTAACGTTATGTCGGCAACCCCGATCTAAAG 11825
 Qy 538 LysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyr 557
 Db 11826 CCGGAAACGGCCACACCTTAAGTCAGCTTACAAATTTAATGGCGATGCGATGCGATTC 11885
 Qy 558 AsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp 577
 Db 11886 TCGACCGGCTGGTATAGCGCAGTACCGATATATCGATGCTGGAAGTGATT----- 11939
 Qy 578 GlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGln 597
 Db 11940 GGCAGCTTTAACCTGATACACCGCAGTAAACGCAATATCTCTAAATTCACCAAC 11999
 Qy 598 SerGlyAlaAspPheTyrGlyValGluGlyGluIleTyrPheLysProThrProArgTyr 617
 Db 12000 GAAGATGCCAATTTATATGGCGCTAAGTTACGGCAGCTCTAC----- 12041
 Qy 618 ArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeu--- 636
 Db 12042 -----CTATTAGCCGATACCCAGCAGTGGTAAGTGGCAGATGCTCGGTAAACTCAAC 12092
 Qy 637 -----ProGlyArgGluAspAlaTyrGlyAsnArgPro-PheIleAla-ClnAspAspG 654
 Db 12093 ATCACCGTGGCGAGCGGATGAA---GGTAATGAACCACTTTATCAATCAACCGGTTA 12149
 Qy 654 InAsn-----AlaProArgValProAlaAlaArgLeuGly-PheHisLeuLys----- 669
 Db 12150 CAACCGAGTTAGCCTCAGCCACCACTCGCGATTCGGAAACCGCTTGGCATGGCAA 12209
 Qy 670 ---AlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGln 688
 Db 12210 TGGGTGGCGACAAAGATCGGCTTATGATCGC----- 12242
 Qy 689 AsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAla 708
 Db 12243 -----CGTTAGAGAACCAACAGCAGCTATTGCTACTGAACTTAAGCAGT 12290
 Qy 709 AsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeu 728
 Db 12291 -----AGCATCAATGGCAAGAGCTGAGTTTAACTTTGCGATCAATAACCTG 12338
 Qy 729 LeuAsnGlnSerValTyrAla----- 735
 Db 12339 TTGTAT---ACTTACTAGCTGCCATCGTGGTGTGAAGATTGCCGATTTAAGGCC 12395
 Qy 736 ---HisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThrGlyGlyVal 754
 Db 12396 GATAGCAGCAATGGCTTTAGCCAAGTCGGGCTCGGCTCGCTCATTTGAATTAGTGCA 12455
 Qy 755 AsnValLysPhe 758

Db 12456 AGCTACCGCTTC 12467
 RESULT 39
 LOCUS AE000694/c
 DEFINITION Aquifex aeolicus VF5 section 26 of 109 of the complete genome.
 ACCESSION AE000694 AE000657
 VERSION AE000694.1 GI:2983162
 KEYWORDS
 SOURCE Aquifex aeolicus VF5
 ORGANISM Aquifex aeolicus VF5
 Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 REFERENCE 1 (bases 1 to 16256)
 AUTHORS Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M., Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
 TITLE The complete genome of the hyperthermophilic bacterium Aquifex aeolicus
 JOURNAL Nature 392 (6674), 353-358 (1998)
 MEDLINE 98196666
 PUBMED 9537320
 REFERENCE 2 (bases 1 to 16256)
 AUTHORS Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L., Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M., Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego, CA 92121
 COMMENT Putative indicates no similarity to known proteins
 Hypothetical indicates similarity to a protein of unknown function.
 FEATURES
 source
 1. .16256
 /organism="Aquifex aeolicus VF5"
 /mol_type="genomic DNA"
 /strain="VF5"
 /db_xref="taxon:224324"
 63. .476
 /gene="spsK"
 /note="synonym: aq_518"
 63. .476
 /gene="spsK"
 /codon_start=1
 /transl_table=11
 /product="spore coat polysaccharide biosynthesis protein SpK"
 /protein_id="AAC06769.1"
 /db_xref="GI:2983168"
 /translation="MNGKAGASSKGGNFVYTIIRKAKAGEPLRVVDDIYMSPTYIDA
 AKELWKLLENKPYGIYHVTNGSYGYSWYFAVKILEYSLKTDIKPVRHTEFKANR
 PLWSPLASRGIGIKLRNWEALKDFINALINLSY"
 445. .1518
 /gene="rfe"
 /note="synonym: aq_519"
 445. .1518
 /gene="rfe"
 /codon_start=1
 /transl_table=11
 /product="undecaprenyl-phosphate-alpha-N-acetylglucosaminyltransferase"
 /protein_id="AAC06768.1"
 /db_xref="GI:2983167"
 /translation="MLFPLITFLISFPLCFLEIKYINLNFHDQKGVQKFFHRRPTPR
 IGLAIYLSVLVSIVILNNKPFSEKRFLEFMSAPPVFPFGLLEDITKKVSPKWRLL
 IGFISGLAPYLISAKVTVDIPFVDSLSISLISLPTAFALAGVSHAFNIDGPNP
 LASGVAMVFGAYAVSVFLHNDLFVLVNLNLLIATLGLFVWNPFLGLFDGGAYF
 LGFCAGAILVNNKYPDIPWPFMLLVIPVWETLFSIYRRKRLKDYPPHMPDALHF
 HOLLYRLLKFTIGSNIEPFKRNSTSPFLWMLICLIPVAVLFDWNTPLILMPTFAF
 IYFTWLYFRIVKFTKPIRIK"
 1519. .2412
 /gene="murB1"
 /note="synonym: aq_520"
 1519. .2412

```

/gene="murBi"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylenolpyruvoylglucosamine reductase"
/protein_id="AAC06766.1"
/db_xref="GI:2983165"
/translation="MFLKNVPLQNLTIKIGRVSFYAEPSDLKEISLCLDFSKSD
ENKVFYGLGNSNTIFGVDGLVNLKAGFKVEIKGFFVEAFSGFLKDLIRFSYK
ENKVFYGLGNSNTIFGVDGLVNLKAGFKVEIKGFFVEAFSGFLKDLIRFSYK
ENKVFYGLGNSNTIFGVDGLVNLKAGFKVEIKGFFVEAFSGFLKDLIRFSYK
YKSPFKLGVIVKVFEPFERSKENILPKYKIRIRKEQPINLPSTGKFRQPEGN
FAGKLEKAGLKGFLKKNVGFSEKHANFLVNGGTTSEVVDLINIAKERVYENFGIV
LREEVKLISSGSDGKWVLGA"
2369..3244
/gene="ddla"
/notes="synonym: aq_521"
2369..3244
/gene="ddla"
/codon_start=1
/transl_table=11
/product="D-alanine:D-alanine ligase"
/protein_id="AAC06765.1"
/db_xref="GI:2983164"
/translation="MKRVVMGGRSREISLTKGAKAVAKALRELGHEVYBLDLKEL
PCKLRIKPKVFIAGHRYGDETVQGLLEILDIPYTGSDTIAGAVSIDKFTKRIY
KSLGINTPMETFIISGDIWNTWKNPFAVPREGSGVLKIVESLEERLEKAYALD
LKTERVMVEEFVEGDMTGVILKGEALFVVEIIPKGVYDYCEYTKMGSEYRILKD
EKLKQLOISLISKISLKDFAIDFRVTKGKIFLEVNTPIGMTLSLPPAAK
EKMDFKLISIIIT"
3255..3905
/gene="aq_522"
3255..3905
/gene="aq_522"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC06775.1"
/db_xref="GI:2983174"
/translation="MRGILGLVPSVILYILSLFMKHPFQIREIKVSGLYVEIE
KIEKIDHALGRLGLVLPESNILEFNEKLNRFQGVENKRFSTEGITTEMFQRKA
ISFVNEKMGIFPWEDEYQKPDFTFIYKEVFPFQKILKLLTQGNISVKY
YKDVVMDIGGRFILLPLEDINEKEIHLMKLLKLPDAKIFDIYKGFILLNEG"
3905..5155
/gene="ftsA"
/notes="synonym: aq_523"
3905..5155
/gene="ftsA"
/codon_start=1
/transl_table=11
/product="cell division protein ftsA"
/protein_id="AAC06770.1"
/db_xref="GI:2983169"
/translation="MKRLAALDIGSOKTVFVIGERDSYGDIIHIGFVPSRGIKGV
INDLSAKGSILRAMKEAEAMAGLVREVVYVNGGTTKNGTVKSONKDTISIKVS
SELESHIQRLHERCLMKAEKGEYIVTAPKYLDDHTVEKNPLGLVSGKLSVEMH
VVKVSTILRLKALREUGNVFVGRVNAIASADSVLTHDEKEDGLDLMGAGLTD
YSYTEGRPIVITGPPFGGINITKUSYMLKIDTETAESVKVNGHGVAFESLVDDEVV
KIRPREDRIPITQKQVABEIIQSRVEEIVKFEIKARGVPLNQINSIGVITGGTA
NLKIRLEIHEMTGLPVRIGLPQIGVLRKEIKENPKYATVCGLLRQAFVSGTIDITES
TNSGTNFFKNLGERFKWPELL"
5168..6271
/gene="ftsZ"
/notes="synonym: aq_525"
5168..6271
/gene="ftsZ"
/codon_start=1
/transl_table=11
/product="cell division protein ftsZ"
/protein_id="AAC06771.1"
/db_xref="GI:2983170"
/translation="WEEFVNPCKIKVIGVGGGGSNAVMRYEDGIEGVELYAINTDVQ
HLSTLVPNKIQIGKVTYRGLAGAKPEVGEAALEDIDIKILELDDTMVFISAGLG
GGTGTAAPVIAKTAKEMGLTVAVALPLPFEPGRMEKALKLEKLESADYIVI
HNDKIKELSNRTLTIKDAFKEVDVLSKAVRGITSIVVTPAVINVDFAVRTLBEGG

```

```

LSIIGSGRGDEKADIAVEKAVTSPLEGGTIGARLLVITMTSEDIYDVIDDEVV
ERHSKVPFABEILIFCAVLPEQDFIRVAIVATDFPEKQVGEKVKFKVKEEK
EPKPEPLSDTITLYEEEPVAVIRKKNKLL"
6268..7032
/gene="moaC"
/notes="synonym: aq_527"
6268..7032
/gene="moaC"
/codon_start=1
/transl_table=11
/product="molybdenum cofactor biosynthesis moaC"
/protein_id="AAC06772.1"
/db_xref="GI:2983171"
/translation="MRTVDITTKITREAKAYGRILRPETVKLIKENKVPKGNLVE
ATKLSIGCAKKTGELLPCHPIDLDVALEKVNEDLEVSTVGRGARTGYEMEAL
TAVTALLNYDMCKALDSDSMVIERVKLESGGKSDMFRDLGVKVNLAHNEGLRK
IADYDKLGATFAEABEILYISIGNLPINKIEIRSLERVISLYDFRRNPKEVKEIRV
GMSDDALIILPESEKIRFFPFTFGIIGNLLCR"
7020..7589
/gene="dedF"
/notes="synonym: aq_528"
7020..7589
/gene="dedF"
/codon_start=1
/transl_table=11
/product="phenylacrylic acid decarboxylase"
/protein_id="AAC06774.1"
/db_xref="GI:2983173"
/translation="MOKIALCITGASGVYIGIKLQMLELDFSDVLVTSRNVKVLK
EHSLTFRSVLKLKNVRIHERNDFTSPASGRVHVGVVVPVPCSTNTLSIANGI
KNKLIHRVGEVALKRVFLVLLVREAPYNEIHLNMLKITRMGGVVVPASPFAFHKPQ
SIDDMINVVKKLLDVLRIEHLNLYKRWG"
7020..7589
/gene="oprC"
/notes="synonym: aq_529"
7020..7589
/gene="oprC"
/codon_start=1
/transl_table=11
/product="outer membrane protein c"
/protein_id="AAC06764.1"
/db_xref="GI:2983163"
/translation="MKRTLLGALFSAFAQVNEEIOVIGKREVLTDVRELPAK
DVGEALVNIPIGVKVKGAJANDVVGIRGKDEVNOLFAGRVVYNACNRMDPGIFH
VDFSEVSEVVIKGFEDVNYGAVGTAVTKKEKMEGRITTVADNWSNTINPSFY
FSYKDRLSFLIGYAFRKPVEDGKRIKILYPSGNFSAISQDEPSTAFNIHTAW
AKLYKITDGVKFKLDYHQRTADVLIPYIMMDGIYDVRVNLGLEGRKFAQLYGS
SVRHWMTNPKRVISQNAQGYNTGYAKSKVYKGFSGFSGFIDFIRYKAAQT
TMYRNMGYRTQNTIPDVVYVNFGLYGRYKLSPLKRLVAGLRIDMSKTKADSGKA
NINLYRHYNTTDSQTDVYPSGNVQVYELKEGIELFAGLSAVRVPDPQRYFALD
RMGMENTMDGVNPKLPDPERNNELDLGILTGERYSGLERTFLSYVKDYIVVRAT
APSSGNMONTKMSYTNIDAVYFGFELSGSVAITDTIFFDGNVAYTRKDDTYPA
KNITDKDIAPLPTARLALYDTGMVFGVGLAATQDNVDSDLQEBKTSYGYIN
LKGGVNRNLRVVAGINNLFDLKLYTHLSYLRNPFRTGVKPEPGRYYLSVYTF"
10030..10818
/gene="aq_531"
10030..10818
/gene="aq_531"
/codon_start=1

```

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0.000109	243.00	16256	174
Percent Similarity:	35.90%	Conservative:	115
Best Local Similarity:	21.61%	Mismatches:	301
Query Match:	6.02%	Indels:	216
DB:	1	Gaps:	35

US-09-936-377-2 (1-758) x AE000694 (1-16256)

Qy 29 GlnSerValGlyIeuGluThrValThrValValGlyIysSerArgProArgalaThrSer 48

Db 9909 CAGGAAGTGAACCTTGAAGAAATTCAAGTTATCGGAAGAGA----- 9868


```

Db      8086  CTTGAAGGTGGGTTAAATTACAGG-----AACCTGAGGTT 8051
QY      723  llyala-----AspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSe 741
Db      8050  TGTGCGGAATTAACAACCTGTTTGAATAACTCTACTATACGACCTCTTTATTGAG 7991
QY      741  rAsp-----ThrProGlnMetGlyArgSerPheThrGlyGlyVa 754
Db      7990  AAATCCCTTCAGAACGGGTGTAAGGTTCCCGAGCCCGAAGGACTTATTATCTTCGTT 7931
QY      754  lAsnValLysPhe 758
Db      7930  TTCGTACACCTTC 7918

RESULT 40
LOCUS   AE016813              336182 bp    DNA        linear    BCT 28-FEB-2003
DEFINITION  Vibrio vulnificus CMCP6 chromosome II section 6 of 6 of the
complete sequence.
ACCESSION AE016813 AE016796
VERSION   AE016813.1 GI:27359304
KEYWORDS
SOURCE    Vibrio vulnificus CMCP6
ORGANISM  Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 336182)
AUTHORS   Rhee,J.H., Kim,S.Y., Chung,S.S., Kim,J.J., Moon,Y.H., Jeong,H. and
Choy,H.E.
TITLE     Complete genome sequence of Vibrio vulnificus CMCP6
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 336182)
AUTHORS   Jeong,H., Moon,Y.H. and Kim,J.J.
TITLE     Direct Submission
JOURNAL   Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong,
Yuseong-gu, Daejeon 305-811, South Korea
REFERENCE 3 (bases 1 to 336182)
AUTHORS   Rhee,J.H., Kim,S.Y., Chung,S.S., Lee,S.E. and Choy,H.E.
TITLE     Direct Submission
JOURNAL   Submitted (13-DEC-2002) Department of Microbiology, Genome Research
Center for Enteropathogenic Bacteria, Chonnam National University
Medical School, Hak-1-Dong, Dong-Gu, Kwang-Ju 501-746, South Korea
FEATURES             Location/Qualifiers
     1..336182
         /organism="Vibrio vulnificus CMCP6"
         /mol_type="genomic DNA"
         /strain="CMCP6"
         /db_xref="taxon:216895"
         /chromosome="II"
     288..1034
         /locus_tag="VV21363"
     288..1034
         /locus_tag="VV21363"
         /note="COG0583"
         /codon_start=1
         /transl_table=11
         /product="Transcriptional regulator"
         /protein_id="AAO08249.1"
         /db_xref="GI:27359305"
         /translation="MAERLCLTPGVSHVKKLVLLGSKVKKKKIETERGQLL
YDFANDLRKHEELNKLQGLVSPVIFSSDCLSPTEELKLEELKELGEGSVIVH
GNNVQTADVLTQVFGDKKTLSEFHHNFQILGRDSDVPIDTVMSPLPEKV
REETHNGYIAETCRWIPMGSRIMQYTHMNGRSSLVWCVDVQNSQFASVSTPLDIQV
RFLMKVTHDGLDITAKTLQNVILIESLFTG"
         /complement(1101..1934)
         /locus_tag="VV21364"
         /complement(1101..1934)
         /note="COG0005"
         /codon_start=1
         /transl_table=11
         /product="Purine nucleoside phosphorylase"

```

```

/protein_id="AAO08250.1"
/db_xref="GI:27359306"
/translation="MTDFVNAQLEIITYSKPNYQPKVAFILGSGSLGVIADELDDKIV
IFYQLEGGPPYSTVVEHSGELVLGKLGIDIVCMKGRGHVYEHQTMKVTNPTFKR
LGCEFLLTNAGSLRPERIDVGLSVFENDHINTWPEPMTGPNDGAPRPFSLNLA
YKSLRSALTVAEAGEIHLNQGVSFVTFNFEETAARIMWQIIGGDVGVMSVVEV
ISAAHCGPLVALCAITNMAEGLSDVTLSHAQTLACAKLAELADFIKLIKSHFAT
S"
2072..2944
/locus_tag="VV21365"
2072..2944
/locus_tag="VV21365"
/note="COG0583"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator"
/protein_id="AAO08251.1"
/db_xref="GI:27359307"
/translation="MLRVFDKVAQIQSFRAEQNLNITKSPLSAQIKLEBEVLGTALF
BNTNRVOLTKAGEOLQVECVTLPAQFERSIHRVKQCARERQQLDIGLMSIIFWAGF
GPAFHALQOLNPHATNLNLELSPEKQVQLLHQIDLGLVRFADTRDVAPLLSECLYQ
ECMVALPSEHPLAQRGLTSLDKSADFWMLKQENSSTBLIAEYCAKAGYRNIVQ
EVVEPNTLIUAVISARQAVSIVPQSYANLSWPQVRFIPLKQRIIPAGVYALYHPEAQSN
KRMIDELKQAIERK"
complement(3014..5149)
/locus_tag="VV21367"
complement(3014..5149)
/note="exoptease Vcc"
/codon_start=1
/transl_table=11
/product="Conserved hypothetical protein"
/protein_id="AAO08252.1"
/db_xref="GI:27359308"
/translation="MTLOQKRSLILALIFGLSGGSGSGSSPTVQDPKDPDPTPT
TWPDRETKSGQCTATOLJADTFSSASQIANADYCMSSWSTSOIEADALFSEONLA
LVNDALLEVNYQSEETAKHYILAFIEXAAYKRNHTYARKIAPFSTELSONIAT
YSAFTLNTHALEMGDEQKALGSMILIVDSVRQLHIAASDVAFALNQFDEQRLTENYY
RELAUKITDDDKLKEKMTQTALILQPPGQSDRVWVGIAEVMNAYAPEHAELG
XKAIFNNIFVAMAGHSGSDDFYQLLNDHRYIDTLTSFVSNQWALGTFDEPLIGNAA
LOPAKRLQRIMTSHDCQGAQIILQMTLAQAEQVCTRLNDKEVDFTVAKTNQO
WPDNDSDVDVIVFKTDYDYSFSLFTNTTNGCFLELRDPSQANNVPRFVAYQN
GWDDPSIILNLEHVVYLDGRFNQYGDPTTMRGHIWVMLEGEAVYVYKEGHHAA
LVLGKKTHTLPEVATSYNDQVNLRYKNGVLGVRLFEKHEPEVQELHLSRLGEYK
QWVYDEQIGPIYADFESLWDDVTKDIDDTVPPEKPEKPEKPEKPEKPEKPEKPEK
EKLFFVDVAANTRELTTLQGGSDADLYACFDKVCYHHDYQWINTFHGNSIQIEKE
SDGYVYKAGRYILISGREAFVRVTAQAQ"
5559..5942
/locus_tag="VV21369"
5559..5942
/locus_tag="VV21369"
/note="COG3094"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein conserved in bacteria"
/protein_id="AAO08253.1"
/db_xref="GI:27359309"
/translation="MMVYLLKKAHLGLIVLSFLSALRAYVYGFKASTWLMKEWPLKV
HKVITVMLISALCLTISQYFEADTWLTKELLLLVAVGVGAMLAFAKQLNRQRTA
FTSVTCLIFAMFYIAKTHVPILS"
complement(5952..7793)
/locus_tag="VV21370"
complement(5952..7793)
/locus_tag="VV21370"
/note="COG0768; Penicillin-binding protein 2"
/codon_start=1
/transl_table=11
/product="Cell division protein FtsI"
/protein_id="AAO08254.1"
/db_xref="GI:27359310"
/translation="MFARAIITAFILACGLVLIANLYLVNHNHFADYQTRSDNRI
KILPIAPVRGLIYDRHGOVLADNLIYVDELIPEBAENIPDTIKMLKRYISLSDQEA
QFYQRLKQTHFRAITIAEDLSEQVARESVHQYQLPGVQISTDFRYPYKASLTHT

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 16:19:56 ; Search time 366 Seconds
(without alignments)
5590.639 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTLKPIVLISILLINTPL.....FLSDTPQMGRSFTGGVNVK 758

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEVS=xlh
-Q=/cgn2.1/USPTO spool/US09936377/runat 14112003 104449 22098/app query.fasta_1.903
-DB=N_Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOEPCI=0
-LOOPEXT=0 -UNITS=bits -SPART=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pct -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09936377 @CGN 1.1.0 @runat 14112003 104449 22098 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4036	100.0	2277	21	DNA encoding a Nei
2	3828	97.3	2295	21	Neisseria meningit
3	3828	97.3	2295	21	Nucleotide sequenc
4	3827	97.3	2277	24	Neisseria meningit
5	3827	97.3	34980	21	Neisseria meningit
6	3827	97.2	2277	21	Neisseria meningit
7	3822	97.2	1437668	21	N. meningitidis B
8	3820	97.1	2277	21	Neisseria meningit
9	3838	95.1	2277	21	Neisseria gonorrhoe
10	3812	94.4	2208	21	Neisseria meningit
11	3812	94.4	22863	21	N. meningitidis pa
12	3706	91.8	2184	25	N. gonorrhoeae nuc
13	1808	44.8	1020	21	Neisseria meningit
14	1791	44.4	1020	21	Neisseria meningit
15	1763	43.7	1020	21	Neisseria meningit
16	1585	39.3	2457	22	Neisseria gonorrhoe
17	1585	39.3	2846	24	Moraxella catarrha
18	1585	39.3	89047	22	M catarrhalis MCAL
19	1205.5	29.9	2942	24	Genomic fragment #
20	1205.5	29.9	269223	22	M catarrhalis MCAL
21	1202	29.8	2742	22	Genomic fragment #
22	803	19.9	522	21	Moraxella catarrha
23	651.5	16.1	2127	24	Neisseria gonorrhoe
24	271.5	6.7	1173	22	M. capsulatus gene
25	230	5.7	1830121	17	Pseudomonas sp hea
26	229.5	5.7	2145	24	Haemophilus influe
27	229.5	5.7	26778	21	Neisseria meningit
28	229.5	5.7	349980	21	N. meningitidis pa
29	229.5	5.7	349980	21	Neisseria meningit
30	227.5	5.6	837096	21	N. meningitidis pa
31	226.5	5.6	2307	24	Nucleotide sequenc
32	223	5.5	2586	22	Pseudomonas sp hea
33	223	5.5	2139	25	N. gonorrhoeae nuc
34	223	5.5	2368	17	N. gonorrhoeae nuc
35	220.5	5.3	2381	22	Neisseria gonorrhoe
36	214	5.3	32160	24	N. gonorrhoeae or N
37	212.5	5.3	4118	17	E. coli CFT073 gen
38	212.5	5.3	2112	21	NTHI hxC + hxCB g
39	212.5	5.3	2112	22	Neisseria meningit
40	212.5	5.3	8367	21	Neisseria meningit
41	212.5	5.3	349980	21	N. meningitidis pa
42	212.5	5.3	1437668	21	Neisseria meningit
43	208.5	5.2	2112	21	Neisseria meningit
44	208.5	5.2	2112	22	N. meningitidis B
45	207	5.1	2600	17	Neisseria meningit

ALIGNMENTS

RESULT 1
AAA75744
ID AAA75744 standard; DNA; 2277 BP.
XX AAA75744;
AC
XX
22-JAN-2001 (first entry)
DE DNA encoding a Neisseria meningitidis BASB082 polypeptide.
KW BASB082; BASB083; BASB091; BASB092; BASB0101; infection; vaccine;
KW gene therapy; upper respiratory tract infection; bacteremia; meningitis;
XX invasive bacterial disease; ss.
OS Neisseria meningitidis.
XX
XX Key Location/Qualifiers
FH

```

FT FT 1..2277
FT /*tag= a
XX /product= "BASB082"
FN WO200055327-A2.
PD 21-SEP-2000.
XX 07-MAR-2000; 2000WO-EF01955.
XX 12-MAR-1999; 99GB-0005815.
XX 21-APR-1999; 99GB-0002094.
XX 23-APR-1999; 99GB-0005503.
XX 28-APR-1999; 99GB-0009787.
XX 07-MAY-1999; 99GB-0010710.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX Defrenne C, Delmelle C, Ruelle J;
XX WPI; 2000-602119/57.
XX P-PSDB; AAB18719.
XX Novel polypeptides designated BASB 082, 083, 091, 092, and 101 derived
XX from meningococcus bacterium useful for producing vaccines against
XX infections and in diagnostic assays -
XX Claim 11; Page 98-99; 108pp; English.
XX The present sequence encodes a BASB082 polypeptide. The specification
XX describes BASB082, BASB083, BASB091, BASB092, and BASB101 polypeptides.
XX The polynucleotides and polypeptides are useful as diagnostic reagents
XX and for diagnosing N. meningitidis infection. The polynucleotides may
XX be used as hybridisation probe for RNA, cDNA and genomic DNA to isolate
XX full-length cDNAs and genomic clones encoding BASB082, BASB083, BASB091,
XX BASB092 or BASB101 polypeptides and to isolate cDNA and genomic clones
XX of other genes that have a high identity particularly high sequence
XX identity to BASB082, BASB083, BASB091, BASB092 or BASB101 genes. The
XX vaccine compositions are useful for inducing an immunological response
XX in humans. The polynucleotides encoding BASB082, BASB083, BASB091,
XX BASB092 or BASB101 polypeptides are useful in gene therapy to induce
XX an immunological response. The polypeptides are useful for treating
XX upper respiratory tract infection, invasive bacterial diseases, such as
XX bacteremia and meningitis.
XX SQ Sequence 2277 BP; 596 A; 713 C; 553 G; 415 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 2277
Score: 4036.00 Matches: 758
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAA75744 (1-2277)
QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
DB 1 ATGGCACAACACTACATCAACCCATGTTTATCATCTTTTATCATCAACACACCCCTC 60
QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
DB 61 CTCGCCCAAGCGCATGAACCTGAGCAATCGTGGGCTTGGAAACGGTTACCGCTCGCGC 120
QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
DB 121 AAAAGCCGTCCGGCGGCACATCAGGGCTGCTGCACACTTCGACCCCTCCGACAAATC 180
QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
DB 181 ATCAGCGCGACACCTTGCACAAAAGCGTCACCTTGGGGGATGCTTTGACGGCGTG 240

```

```

QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
DB 241 CCGGGCATTACGCTCGCAATACGGCGCGCGCGCTCCGCTCCGTTATTTCGGGGTCAA 300
QY 101 ThrGlyArgArgIleIleValLeuAsnHisIleGlyGluThrGlyAspMetAlaAspPhe 120
DB 301 ACAGGACAGCGGATTAAAGTATTGAACCATCACGGCGAAACAGGCGATATGGCGATTTT 360
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValIleLeuArg 140
DB 361 TCGCCGATCACGCCATTATGTTAGTACCGCTTGTGCAACAGTCCGAAATCTTGGC 420
QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
DB 421 GGGCCGGTTACGCTCTTGTACAGCTCGGGCAATGTGGCGGGGCTGGTGCATGTTGCCGAT 480
QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
DB 481 GGCAAAATCCCCGAAAAAATGCTGAAAACGGCGTATCGGGCGAACTCGGATTGGCTTTG 540
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
DB 541 AGCAGCGGCAATCTGAAAAAATCTCAGCTCCGGCGGCATCAATATCGGTTTGGGCAAAAAC 600
QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
DB 601 TTTGTTATTCACACGGAAGGGCTGTACCGCAATCGGGGATTACCGCGTACCGCGTTAC 660
QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
DB 661 CGCAATCTGAAACGGCTCGCCGACAGCCACCGCATTCGCAACGGCGGCATCGGGCT 720
QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr 260
DB 721 GTCTTGGTGGCGAAAAAGGTTTATCGCGCTAGCTACAGCAACGTCGCGCAACATAT 780
QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280
DB 781 GGTCCTGCTGCCACAGCCAGCAATACGATGATTCGCCACGCGACATCATCTGGCAAAAG 840
QY 281 SerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVal 300
DB 841 AGCTTGATTAACAAACGCTATTATCAGCTTTATCGCCTGTTGACCGAAGAACGCTC 900
QY 301 AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
DB 901 GATTACGACAATCCGGGCTTGAGCTGGCGCTTCCACGACGACGATGTCACACGCCCAT 960
QY 321 AlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGlu 340
DB 961 GCCCACAACGGCAACCTTGGATAGACCTCGCCCAACAAACGCTACGAACTCCGCGCGGAA 1020
QY 341 TrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
DB 1021 TGGAAAGCAGCATTCGCCGCTTTTGAAGCCCTTCGCGCTACACCTGAACCGCAACGACTAC 1080
QY 361 HisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsn 380
DB 1081 CACACGACGAAAAACGCGCATGCGTGCAGAAACTTTTAAACACCAACCGCAAAAC 1140
QY 381 AlaArgIleLeuLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGln 400
DB 1141 GCCCGCATCGAGTTGGCCACCAACCCATAGGCGCTCTGAAAGGCGAGCTGGGGCGTGC 1200
QY 401 TyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMet 420
DB 1201 TATTGGGCAAAAATCCAGTGTCTTATTCGCACATCCGAAAGCGGTCAAAACCGATG 1260
QY 421 LeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAsp 440
DB 1261 CTGCTTGACATAAAGTGCAACATTACAGCTTTTTCGGTGTAGAACAGCGCAAACTGGGAC 1320
QY 441 AsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAsp 460

```

```

Db 1321 AACTTCAGCTTGAAGCGCGCTACGCTGGAAGAAACAAAAGCTTCATCGCTACGAC 1380
Qy 461 LysAlaLeuLeuAspArgGluAenTyrTyrLysGlnProLeuProAspLeuGlyAlaHis 480
Db 1381 AAGCATTTGATTCGGGAAACTACTAAGCAGCCCTGCGCCAGCTCGCGCGGCAC 1440
Qy 481 ArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLys 500
Db 1441 CGCCAAACCGCCGCTGCTTCGACATTCGCGCACTTCGCGCACTGGATTTTCACGCGCAACAAA 1500
Qy 501 LeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGlnLeuTyrAlaHis 520
Db 1501 CTCAGCTGACCGCTCCCATCAGGAAGCGCTGCGCTCAACGCAAGAGCTGTACGCACAC 1560
Qy 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
Db 1561 GGCAAAACAGCTTGCACCAACACTTTTGAAGTCGCGCAACAACTGGAACAAAGAGCT 1620
Qy 541 SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAla 560
Db 1621 TCCAAACAATCGAACTCGGCTTGGCTACGAAGCGCGCTGGCAATACAATCTGGCA 1680
Qy 561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
Db 1681 CTCCTACCGCAACCGCTTCGCGCACTCATTTACGCCCAACCTTAACGACGCGCGCGC 1740
Qy 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAla 600
Db 1741 CCCAAATCCATCGAAGACGACAGCGAAATGAAGCTGCTGCGCTACAACCAATCCGCTGGC 1800
Qy 601 AspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGly 620
Db 1801 GACTTCTACCGCGCGGAAGCGAAATCTACTTCAACCGGACCGCGCTACCGCATCGGC 1860
Qy 621 ValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGlu 640
Db 1861 GTTTCGCGGACTATGTACGAGGCGCTCGAAACCTGCGCTCCCTACCGCGCAGCGAA 1920
Qy 641 AspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPro 660
Db 1921 GATGCTTACGCGCAACCGCTCTTTCATCGCGCAGGACGACCAAAACGCGCTCGCTCCG 1980
Qy 661 AlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeu 680
Db 1981 GCTGCGGCGCTCGCTTCCACTCGAAAGCTCGCTGACCGCGCATCGATGCAATTTG 2040
Qy 681 AspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
Db 2041 GACTACTACCGCGCTGTTTGGCCCAAAACAACTCGCGCTACGAAACGCGCAGCGCGGA 2100
Qy 701 HisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAsn 720
Db 2101 CACCATATGCTCACTTCGCGCGCAACTACCGCGCAATACCGCGCTATGCGAGTGGAT 2160
Qy 721 TrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLeu 740
Db 2161 TGGTACGTCGCAAGCGCAACCTGCTCAACCAATTCGCTTTACGCGCCACAGCAGCTTCTC 2220
Qy 741 SerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2221 TCTGATACGCCACAAATAGGCGCGCGAGCTTTACCGGTGGCTTACAGTGAAGTTT 2274

RESULT 2
AAZ53319
ID AAZ53319 standard; DNA; 2295 BP.
XX
AC AAZ53319;
XX
XX
XX 21-MAR-2000 (first entry)
XX
XX Neisseria meningitidis ORF 147 partial DNA sequence SEQ ID NO:587.
DE
XX

```

```

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR P-PSDB; AAY74557.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
PT
PS Claim 7; Page 414; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 2295 BP; 606 A; 724 C; 553 G; 412 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0 Length: 2295
Score: 3928.00 Matches: 744
Percent Similarity: 98.68% Conservative: 5
Best Local Similarity: 98.02% Mismatches: 9
Query Match: 97.32% Indels: 2
DB: 21 Gaps: 0
US-09-936-377-2 (1-758) x AAZ53319 (1-2295)
QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuLeuLeu 20
Db 19 ATGGCACAACACTACACTCAACCCCATTTGTTTATCAATTTCTTTTAAACAACACACCCCTC 78
QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
Db 79 CTCTCCCAAGCGCATGGAACTGAGCAATCTAGTGGCTTGGAAACGGTCAGCTGCTGGC 138
QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db 139 AAAAGCGCTCGCGCGCCACTTCGGGGCTGTCGACACTTCTACCGCTTCGCAAAATC 198

```

```

QY 115 SerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
DB 199 ATCAGCGGCGACACCTTGGCAAAAGCCGTCACCTTGGGTGATGCTTTAGACGGCGTA 258
QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
DB 259 CCGGGCATTCATGCTCGCAATACCGCGCGCGCATCCGCTCCGCTTATTTGGGGTCAA 318
QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
DB 319 ACAGGCGACGCGATTAAAGTGTGAACCATCAGCGCGAAACGGGCGACATGCGCGCACTTC 378
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArg 140
DB 379 TCTCCAGACCATGCAATCATGTGTGACAGCTCGGCAATGTGGCGGCTGTCGAATCCTGCGC 438
QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
DB 439 GGTCCGGTTACGCTCTTGTACAGCTCGGCAATGTGGCGGCTGTCGATGTTGCCGAT 498
QY 161 GlyIleIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
DB 499 GGCAGAAATCCCGAAATAATGCTGAAACGCGGTATCGGGCAACTCGGAATTCGGTTTG 558
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
DB 559 AGCAGCGGCAATCTGGAAATACTCACGTCGGCGGCGCATCAATATCGGTTTGGGCAAAAC 618
QY 201 PheValLeuHisThrCluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
DB 619 TTTGTATTGCACAGGAAGGCTGTACCGCAATCGGGGATTACCGCGTACCGGGTTAC 678
QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
DB 679 CGCAATCTGAAACCGCTGCGCGACAG-CCACGCGGATTTCGCAACCGGCGAGCATCGGCT 737
QY 241 ValLeuGlyTrpArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr 260
DB 738 GTCTTGGGTTGGGAAAGGCTTTATCGCGCGACGATACAGCGACCGCTCGCGCAATA 797
QY 260 rGlyLeuProAlaHisSerHisGlnTyrAspCysHisAlaAspIleIleTrpGlnLys 280
DB 798 TGGTCTGCTGCCCAAGCAGCAATACGATGATTGCCAGCGCGACATCATCTGSCAAA 857
QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLysPva 300
DB 858 GAGTTTGTATTAAACAAACGCTATTTCAGCTTTATCGGACCTTTGACCGAAGAGACAT 917
QY 300 lAspTyrAspAsnProGlyLysSerCysGlyPheHisAspAspAspAlaHisAlaHi 320
DB 918 CGATTACGACAAATCCGGGCTTGAAGCTTTCAGCGACGACGATGATGCACACGCCCA 977
QY 320 sAlaHisAsnGlyLysProTrpIleAspLeuArgHisLysArgTyrGluLeuArgAlaGl 340
DB 978 TGCCCAACACGGCAACCTTGGATAGACTCGCGCAACAAACGCTACGAACTCCGCGCGGA 1037
QY 340 uTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360
DB 1038 ATGGAAGCAACCGTTCCTCCGGTTTGAAGCCCTTGGCGGTACACCTGMAACCGCAACGACTA 1097
QY 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsnGlnThrGlnAs 380
DB 1098 CCGCCACGACGAAAGAGCGGCGATGAGTAGAACTTTTAAACCAACCAACCAACAA 1157
QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGl 400
DB 1158 CGCCCGTATCGAGTTGCGCACCAACCCATAGCGCTCTGAAAGGCGAGCTGGGGGTGCA 1217
QY 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
DB 1218 ATATTTGGGCAAAATTCAGTGTCTTTATCTGCACATCCGAAGCGGTCAAACACCGAT 1277

```

```

QY 420 tLeuLeuAspLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAs 440
DB 1278 GCTGCTGCACATAAAGTGCAACATTACAGCTTTTTCGGTGTAGAACAGGCAACTGGGA 1337
QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
DB 1338 CAACCTTCAGCTTGAAGCGGGGTACGCTGGAAAAACAAAAAGCCTCCATCCGTACGA 1397
QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
DB 1398 CAAAGCATTTGATTGATCGGGAAACTACTACAACCATCCCTGCGCGACCTCGCGCGCA 1457
QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLy 500
DB 1458 CGGCCAAACCGCGCTCATTCGCACTTTCGGCACTTGGTATTTTCAACGCCACACACAA 1517
QY 500 sLeuSerLeuThrAlaSerHisGlnArgLeuProSerThrGlnGluLeuTyrAlaHi 520
DB 1518 ACTCAGCCTTACCGCTCCCATCAGAAACGCTGCGCTCAACGCAAGAGCTGTACGCACA 1577
QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
DB 1578 CGGCAACACGTCGCCACCAACCTTTGAAGTCGGCAACAAACCTCAACAAAGAGCG 1637
QY 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAl 560
DB 1638 TTTCAACAATATCGAACTCGCGCTGCGCTACGAGCGACCGCTGCGCAATACAATCTGCGC 1697
QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGl 580
DB 1698 ACTCTACCGCAACCGCTTCGGCACTCATCTTACGCCCAACCTTAAACGCGACGCGCG 1757
QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
DB 1758 CCCCAAATCCATCGAAGACGACAGCAAAATGAAGCTCGTGGCTTACAACCAATCCGCTGC 1817
QY 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGl 620
DB 1818 GGAATCTTACCGCGCGGAAGGCAATCTACTTCAACCGACACCGCGCTACCGCATCGG 1877
QY 620 yValSerGlyAspTyrValArgGlyValGluLysAsnLeuProSerLeuProGlyArgGl 640
DB 1878 CGTTTCCGCGCACTATGTACGAGGCGCTCTGAAAAACCTGCTTCCCTACCGCGCAGGA 1937
QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValer 660
DB 1938 AGACGCTTACGGCAACCGGCCACTCATTTGCCCAAGCCGACCAACGCCCCCTCGCGTTC 1997
QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
DB 1998 GGCTGCGCGCTCGGCGTCCACTCGAAAGCCTCGTGACCGACCGCATCGATGCCAATT 2057
QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl 700
DB 2058 GGAATCTACTCCGCGTGTTCGCCCAACAACTTCGCCCGCTACGAAACGCGCACCGCCGG 2117
QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyLysTrpAs 720
DB 2118 ACACCATATGCTCAACCTCGGCGCAACTACCGCGCAATACGCGCTATGCGAGTGGNA 2177
QY 720 nTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
DB 2178 TTGGTTCGTCAAAGCGCAACCTGCTCAACCAATCCGTTTACGCCCAACGACGAGCTTCT 2237
QY 740 uSerAspThrProGlnMetGlyVarSerPheThrGlyGlyValAsnValLysPhe 758
DB 2238 CTCTGATACGCCCAATTCGCCCGAGCTTTACCGCGCGGTGAACGTGAAGTTT 2292

```

RESULT 3

ABQ78299

ID ABQ78299 standard; DNA; 2295 BP.

XX

AC ABQ78299;


```

Db      1278  GCTGCTTGACAAATAAAGTGCACATACAGCTTTTTCGGTGTAGACAGGCAAACTGGGA 1337
Qy      440  pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db      1338  CAACCTTCACGCTTGAGAGCGCGTACGCTGGGAAAAACAAGAAGCTTCATCCGTACGA 1397
Qy      460  pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
Db      1398  CAAAGCATTGATTGATCGGGAATACTACAAACCACTCCCTCCCGAGCTCGGCGCGCA 1457
Qy      480  sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisIly 500
Db      1458  CCGCAAAACCGCGCTCATTCGCACCTTCGGGCAACTGTGATTTTCAACGCCACACAA 1517
Qy      500  sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db      1518  ACTCAGCCTGACGCTCCCATCAGGAACGCTCGCTCAACGCAAGAGCTGTACGCACA 1577
Qy      520  sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
Db      1578  CGGCAAAACGCTGCCACCAACCTTTGAAGTCGGCAACAACACCTCAACAAGAGCG 1637
Qy      540  gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
Db      1638  TTCCAAATATCATCGCTCGCTGGCTACGAAGCGACCGCTGGCAATACAAATCTGGC 1697
Qy      560  aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGl 580
Db      1698  ACTTACCGCAACCGCTTCGGCAACTACATTTAGCGCCAAACCTTAAACGACGAGCGG 1757
Qy      580  yProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db      1758  CCCCAAATCCATCGACGACGACGGAATGAAAGCTCGTGGCTACAAACCAATCCGGTGC 1817
Qy      600  aAspPheTyrClyValGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGl 620
Db      1818  GCACTTCACGGCGGAGGAGGAAATCTACTTCAACCGCACCGCTGATCCGCAATCGG 1877
Qy      620  yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGl 640
Db      1878  CGTTTCGGCGGACTATGTACGAGCGCTCTGAAAAACCTGCTTCCCTACCGCGAGGGA 1937
Qy      640  uAspAlaTyrClyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
Db      1938  AGACGCTACGGCAACCGCCCACTCATGTGCCAGCGCAACCAACGCGCTCGCGTTCC 1997
Qy      660  oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
Db      1998  GGCTGCGCGCTCGCGCTCCACCTGAAAGCTCGCTGACCGACCGCATCGATGCCAATTT 2057
Qy      680  uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl 700
Db      2058  GGACTACTACCGCTGTTCGCCCAAAACAACTCGCCCGCTACGAAACGCGACGCGCGG 2117
Qy      700  yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrpAs 720
Db      2118  ACACCATATGCTCAACTCGCGGCAAACTACCGCGCAATACCGCTGTATGGCGAGTGGAA 2177
Qy      720  nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
Db      2178  TTGGTACGTCAAGCCGACACCTCTCAACCAATCCGTTTACGCCCAACAGCGACTTCCT 2237
Qy      740  uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db      2238  CTCGTATACGCGCAAAATGGCGCGAGCTTTACCGCGCGGTGAAACGTGAAGTTT 2292

```

RESULT 4

AAS97296

ID AAS97296 standard; DNA; 2277 BP.

XX

AC

AAS97296;

XX

```

DT      12-MAR-2002 (first entry)
XX      Neisseria meningitidis virulence gene #101.
XX      Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
XX      infection; Gram-negative bacteria; antimicrobial; ds.
XX      Neisseria meningitidis.
XX      WO200185772-A2.
XX      15-NOV-2001.
XX      08-MAY-2001; 2001WO-GB02003.
XX      08-MAY-2000; 2000GB-0011108.
XX      (MICR-) MICROSCIENCE LTD.
XX      Tang C;
XX      WPI; 2002-066593/09.
XX      P-PSDB; AAU73011.
XX      New peptide encoded by operon including virulence genes of Neisseria
XX      meningitidis, useful as vaccine component for treating or preventing
XX      meningitis and for identifying antimicrobial drug.
XX      Claim 1; Page 408-412; 423pp; English.
XX      The invention relates to a peptide (I) encoded by an operon (II) of
XX      Neisseria meningitidis including virulence genes, or a related molecule
XX      having a 40% sequence similarity at the peptide or nucleotide level in a
XX      Gram-negative bacterium, or its functional fragment, for therapeutic or
XX      diagnostic use. (I) and (II) are useful in the manufacture of a
XX      medicament for treating or preventing a condition (e.g., meningitis)
XX      associated with infection by Neisseria or Gram-negative bacteria. The
XX      product is useful for veterinary treatment and in a screening assay for
XX      the identification of an antimicrobial drug. The vaccines have
XX      prophylactic applications. AAS97196-AAS97305 represent N. meningitidis
XX      virulence genes and related PCR primers of the invention.
XX      Sequence 2277 BP; 600 A; 712 C; 546 G; 419 T; 0 other;

```

Alignment Scores:

```

Pred. No.: 0 Length: 2277
Score: 3927.00 Matches: 743
Percent Similarity: 98.81% Conservative: 7
Best Local Similarity: 97.89% Mismatches: 8
Query Match: 97.30% Indels: 2
DB: 24 Gaps: 0

```

US-09-936-377-2 (1-758) x AAS97296 (1-2277)

```

Qy      1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db      1 ATGGCAACAACACTACACACCAACCACTGTTTATCAATCTTTTATCAACACACCCCTC 60
Qy      21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
Db      61 CTCGCCCAAGCGCATGAATGAGCAATCGTGGATTTGAAACGGTACGCTGCTGCGC 120
Qy      41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db      121 AAAAGCGCTCCGCGCGCACGTCGGGCTGTTGCACACTTCGACCGCTCCGACAAATC 180
Qy      61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db      181 ATCTCCCGCGATACCTTCGCCCAAAAGCGCTCAACTTGGCGGACCTTTTAGACGCGTA 240
Qy      81 ProGlyIleHisAlaSerGlnTyrGlyGlyValAlaSerAlaProValIleArgGlyGln 100
Db      241 CCGGGCATCCACGCTTCGCAATACGGCGGGCGGCGCTCTGCTCCCTCATTCGGGTCAA 300

```

QY 101 ThrGlyArgArgIleLysValLeuAsnHis sli sgl yGluThrGlyAspMetAlaAspPhe 120
 DB 301 ACAGCGAGCGGATTAAGTGTGAACCATCACGCGGAAACAGGCGGATATGGCGGATTTT 360
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 DB 361 TCGCCCGATCAGCCCATATGTAGATACCGCTTGTGCGAACAGGTGCGAAATCCTCGGC 420
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyLeuValAlaGlyLeuValAspValAlaAsp 160
 DB 421 GCGCGCGTGTACGCTCTGTACAGCTCGGCGCAATGGCGGGCTGGTGGATGTTGGCGGAT 480
 QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 DB 481 GGCAGAAATCCCGAATAATGCTGAAACGCGATTCGGCGAACTCGGATTCGGTTG 540
 QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 DB 541 AGCAGCGGCAATCTGGAAAACTCACGTCGCGCGCATCAATATCGGTTTGGGCAAAAAC 600
 QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
 DB 601 TTGTATTGCACACGGAAGGCTGTACCGCAATCGGGGATTACGCGGTACCGGTTAC 660
 QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 DB 661 CGCAATCTGAAACGCGCTGCCGACAG-CACACGCCGATTTGCAAAACGGCGAGCATCGGCT 719
 QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr 260
 DB 720 GTCTTGGTGGCAAAAAGTTTATCGCGTAGCTACGAGCGCTCGGACCAATA 779
 QY 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280
 DB 780 TGGTCTGCTGCCACAGCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 839
 QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
 DB 840 GAGCTTGATTAACAAACGCTATTACAGCTTTATCGCACCTGTGACCGAGAGAGCAT 899
 QY 300 lAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
 DB 900 CGATTACGACAAATCCGGGCTTGAGCTGCGGCTTCCACGACGACGATATGACACGACACA 959
 QY 320 sAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGly 340
 DB 960 CACCCACAGCGGACCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1019
 QY 340 uTrrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
 DB 1020 ATGGAAGCAACCGTTCCCGGTTTGAAGCCCTTGGCGGTACACCTGAAACCGCAACGACTA 1079
 QY 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 DB 1080 CCGCCACGACGAAAAGAGGCGATGCGATGCGAATCTTTTAAACCAACCGCAAAA 1139
 QY 380 nAlaArgIleGluLeuArgHisGlnProIleArgLeuLysGlySerTyrGlyValGly 400
 DB 1140 CGCCCGCATCGAGTTGCGGCACCAACCCCATAGTGTCTGAAAGGCGAGTGGGCGTGC 1199
 QY 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMet 420
 DB 1200 ATATTACAAACAAATCCAGTCTTATCTGCCATATCCGAAGCGGTAAACACCCGAT 1259
 QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTyrAs 440
 DB 1260 GCTGCTTGACAAACAAAGTGCACATACAGCTTTTTCGGTGTAGAACAGGCAAACTGGGA 1319
 QY 440 pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 DB 1320 CAACTTCACGCTTGAAGAGGCGTACGCGTGGAAAAACAAAGGCTCCATTCAGTACGA 1379

QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHis 480
 DB 1380 CAAAGCATTTGATTGATCGGAAAACTACTCAACACCCCTGCGGACCTCGGCGGCA 1439
 QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLys 500
 DB 1440 CGCCCAAAACCGCGCTCATTCGCACTTTCGGCAACTTGGTAATTTACGCGCACACAA 1499
 QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis 520
 DB 1500 ACTAGCCTGACCGCTCCCATCAGAACGCTGCGGTCAACGCAAGAGCTGTACGACAA 1559
 QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
 DB 1560 CGGCAAAACAGCTGCGCACCAACACTTTGAAGTCGGCAACAAACACTCAACAAAGAGCG 1619
 QY 540 sSerAsnAsnIleGluLeuAlaLeuGlyTyrGluLysAspArgTyrGlnTyrAsnLeuAl 560
 DB 1620 TTCCCAACAATATCGAATCGCGCTGGCTACGAGCGGACCGCTGGCAATTAACATCTGCG 1679
 QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
 DB 1680 ACTCTACGCAACCGCTTCGGTAACATATTACGCCCAAACTTTAAACGACGAGCGCG 1739
 QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 DB 1740 CCCCAAATCCATCGAAGACGACAGCAAAATGAAGCTCGTGGCTACAAACCAATCCGCGCG 1799
 QY 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGly 620
 DB 1800 CGACTTCTACGCGCGGAGGCGAATCTACTTCAAAACGACACCGCTACCGCATCGG 1859
 QY 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGly 640
 DB 1860 CGTTTCCGCGCATGTGTACGAGGCGCTGTGAAAAACCTGCTTCCCTACCGCGCAGAGA 1919
 QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 DB 1920 AGATGCTACGCAACCGCTCTTTCATCGCACAGACGACCAAAATGCCCCCGGTTC 1979
 QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
 DB 1980 GGTGCGCGCTCGGCTTCCACTGAAAGCTTCGCTGACCGACCGTATCGATGCCAATTT 2039
 QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
 DB 2040 GGACTACTACGCGGTGTCGCCCAAAACAACTCGCCCGCTACGAAACGCGCACCGCCGG 2099
 QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAs 720
 DB 2100 ACACCATATGCTCAACCTCGCGCGCAAACTACCGCGCAATACGCGTATGCGAGTGGAA 2159
 QY 720 nTyrTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheIle 740
 DB 2160 TTGGTACGTCAAGCGCAACCTGCTCAACCAATCCGTTTACGCCCAACAGCAGCTTTCT 2219
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 DB 2220 CTCTGATACGCGCAAAATGGCGCGAGCTTTTACCGCGCGGTGAACGTGAAGTTT 2274

RESULT 5

AAF21609/c

ID AAF21609 standard; DNA; 349980 BP.

XX AAF21609;

AC AAF21609;

DT 13-MAR-2001 (first entry)

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:110.

DE Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;

KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;

KW ds.

XX OS *Neisseria meningitidis*.
 XX PN WO200066791-A1.
 XX XX 09-NOV-2000.
 XX PF 08-MAR-2000; 2000WO-US05928.
 XX PR 30-APR-1999; 99US-0132068.
 XX PR 08-OCT-1999; 99WO-US23573.
 XX PR 28-FEB-2000; 2000GB-0004695.
 XX PA (CHIR) CHIRON CORP.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
 XX PI Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;
 XX PI Frazer CM, Grandi G;
 XX XX MPI; 2000-647603/62.
 XX XX *Neisseria meningitidis* B full length genome sequence and open reading
 XX PT frames are used to detect, treat and prevent *Neisseria* infections -
 XX XX Claim 7; Appendix A; 692pp; English.
 XX CC The present invention describes the full length genome of
 XX CC *Neisseria meningitidis* B (NMB). The sequences in AAF21544 and AAF21607
 XX CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 XX CC sequence was too long to go in a record on its own it was split into 8
 XX CC sequences which overlap each other at the beginning and end of each
 XX CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 XX CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 XX CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 XX CC *Neisseria* proteins given in AAF21589 to AAF21593, and AAF21589 to
 XX CC AAF21606 represent PCR primers which are used in the exemplification of
 XX CC the present invention. The NMB genome and fragments from it have
 XX CC antibacterial activity, and can be used in vaccines and gene therapy.
 XX CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the
 XX CC proteins can be used in compositions for treating or preventing infection
 XX CC due to *Neisseria* bacteria or as a diagnostic reagent for detecting the
 XX CC presence of *Neisseria* bacteria or of antibodies raised to *Neisseria*
 XX CC bacteria. Computers, computer memory, computer storage medium or computer
 XX CC databases can be used in a search to identify open reading frames (ORFs)
 XX CC or coding sequences within the NMB genome. The DNA sequences provide
 XX CC further opportunities to find antigenic or immunogenic proteins which are
 XX CC more effective in vaccines than the outer membrane proteins currently
 XX CC used.

SQ Sequence 349980 BP; 81351 A; 86755 C; 95584 G; 86290 T; 0 other;

Alignment Scores:

Pred. No.:	0	Length:	349980
Score:	3927.00	Matches:	743
Percent Similarity:	98.81%	Conservative:	7
Best Local Similarity:	97.89%	Mismatches:	8
Query Match:	97.30%	Indels:	2
DB:	21	Gaps:	0

US-09-936-377-2 (1-758) x AAF21609 (1-349980)

QY 1 *MetaAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu* 20
 Db 85503 ATGGCAAAACACTCACTCAAAACCAATGTTTATCAATCTTTAATCAACACACCCCTC 85444

QY 21 *LeuAlaGlnAlaHisGlnThrGluGlnSerValGlyLeuGlnValThrValValGly* 40
 Db 85443 CTGCCCCAGCGATGAATGAGCAATCGGTGGATTTGAACCGCTGAGCTGCTGCGC 85384

QY 41 *LysSerArgProAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle* 60
 Db 85383 AAAAGCCGTCCGGCGCGCAGTCCGGGCTGTTGCACACTTCGACCGCCTCCGACAAAATC 85324

QY 61 *IleSerGlyAspThrLeuArgGlnIleValAlaValAsnLeuGlyAspAlaLeuAspGlyVal* 80
 Db 85323 ATCTCCGGCGATACCTTGGCCAAAGAGCGTCAACTTGGGCGAGCGCTTTAGACGCGCTA 85264

QY 81 *ProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAlaProValIleArgGlyGln* 100
 Db 85263 CCGGGCATCCAGCTTCGCAATACGGCGCGCGGTCTGCTCCCGTCATTCCGGGTCAA 85204

QY 101 *ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe* 120
 Db 85203 ACAGGCGAGCGGATTAAGTGTGAACCATCAAGCGGAAACAGGCGATATGCGCGATT 85144

QY 121 *SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg* 140
 Db 85143 TCSCCGCATACGCCATTATGTGTAGATACCGCTTGTGCAACAGGTGAAATCTCGCGC 85084

QY 141 *GlyProValThrLeuLeuThrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp* 160
 Db 85083 GGGCGGTTACGCTCTTGTACAGCTCGGCGCAATGTGGGGGCTGCTCGATGTTCCGAT 85024

QY 161 *GlyIleProGlnLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu* 180
 Db 85023 GGCATAATCCCGAAAAATGCTGAAACGGCGTATCGGCGCAACTCGGATTCGCTTTG 84964

QY 181 *SerSerGlyAsnLeuGlyLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn* 200
 Db 84963 AGCAGCGGCAATCTGGAAAACTCAGTCCGGCGGATCAATATCGTTCGGCAAAAAC 84904

QY 201 *PheValLeuHisThrGluGlyLeuThrArgLysSerGlyAspThrAlaValProArgTyr* 220
 Db 84903 TTTGTATTGCACACGGAAGGCTGTACCGCAATCGGGGATTTACCGCGTACCGGTTAC 84844

QY 221 *ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla* 240
 Db 84843 CGCAATCTGAAACGCTCTGCCGACG - CCACGCGGATTCGCAACGGGCGAGCATCGGCT 84785

QY 241 *ValLeuGlyTrpArgLysArgPheTyrArgArgThr - TyrSerAspArgArgAspGlnTyr* 260
 Db 84784 GTCTTGGTGGCGAAAAAGTTTATCGCGGTAGCGTACAGCGACCGTTCGGCAATA 84725

QY 260 *ArgLeuProAlaHisSerHisGlyLysThrAspAspCysHisAlaAspIleIleTyrGln* 280
 Db 84724 TGCTCTGCTGCCACAGCCAGCATCATGATGTCACGCGCGCATCATCTTCGCAAAA 84665

QY 280 *SerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluAspVal* 300
 Db 84664 GAGCTTGATTAAACAAACGCTATTATACAGCTTTATCCGCACTGTTGACCGAAGAGAT 84605

QY 300 *LaspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis* 320
 Db 84604 CGATTACGACATCCGGGCTTGAGCTCGGCTTCCAGCAGCATATATGCAACGCCA 84545

QY 320 *AlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG* 340
 Db 84544 CACCCACAGCGCGAGACCGGTGATAGACCTGCGCAACAAACGCTACGAACTCGTGCCA 84485

QY 340 *uTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr* 360
 Db 84484 ATGAAGCAACCGTTCCTCCCGTTTGAAGCCCTTCGCGCTACCTGAAACCGCAAGACTA 84425

QY 360 *rHisHisAspGlyLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs* 380
 Db 84424 CCGCCACGAGAAAAAGCAGCGGATCGATCGCAAAACTTTTTTAAACAAACCAAAA 84365

QY 380 *naIaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValG* 400
 Db 84364 CGCCCGCATCGATGTCGCCACCAACCCATAGTCTGTGAAGAGCGGCTGGGGGCTGCA 84305

QY 400 *nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe* 420
 Db 84304 ATATTTACAAAAAATCCAGTGTCTTTATCTGCCATATCCGAGCGGTTAACACCGAT 84245

QY 420 tLeuLeuaspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
 DB 84244 GCTGCTTGACAAAGTGCACATATACAGCTTTTTCGGTGTAGAACAGGAACTGGGA 84185
 QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 DB 84184 CAACTTCAGCTTGAGAGGGGTACGGGTGGAAACAAACAAAGCCCTCATTCAGTACGA 84125
 QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
 DB 84124 CAAAGCATTTGATTCGGGAAATCTACTACAAACCCCTCGCCGAGCTTCGGCGGCA 84065
 QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLy 500
 DB 84064 CGCCAAACCGCCGCTCATTCGCACCTTCGGGCACTGGTATTTACGCCACACACAA 84005
 QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
 DB 84004 ACTCAGCTTCAGCGCTCCCATCAGGAACGCTCGCTCAACGCAAGAGCTGTACGCACA 83945
 QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
 DB 83944 CGGCAACACGCTCGCCACCAACCTTTGAGTTCGGCAACACCTCAACAAAGAGCG 83885
 QY 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAl 560
 DB 83884 TTCCACAAATATCGAACTCGCGCTGCTACGAAGCGCACCGCTGGCAATCAATCTGGC 83825
 QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGl 580
 DB 83824 ACTTACCGCAACCGCTTCGGTAACTACATTTACGCCCAACCTTAAACGAGCGAGCGG 83765
 QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 DB 83764 CCCAAATCCATCGAAGACGACGAGAAATGAAGCTCGTGGCTACCAACCAATCCGGCGC 83705
 QY 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgGleGl 620
 DB 83704 CGACTTCTACGGCGGAGAGGCGAAATCTACTTCAACCGACACCGCGCTACCGCATCGG 83645
 QY 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGl 640
 DB 83644 CGTTTCGGCGAGATATGTACAGGCGCTGTGAAACCTTCCTCCCTACCGCGGACAGA 83585
 QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 DB 83584 AGATGCTTACGGCAACCGTCCCTTCATCGCAGGACGACCAAAATGCCCCCGGTTCCTC 83525
 QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgGleAspAlaAsnLe 680
 DB 83524 GGCTGCGCGCTCGGCTTCACCTGAAGCCTCGCTGACCGACCGCTATCGATGCCAATTT 83465
 QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGl 700
 DB 83464 GGACTTACTACCGGTGTTCGCCCAAAACAACTCGCCGCTACGAAACGGCGACCGCCGG 83405
 QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAs 720
 DB 83404 ACACCATATGCTCAACCTCGCGCAACATACCGCGCAATACCGCTATGCGAGTGGAA 83345
 QY 720 nTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 DB 83344 TTGTGAGTCAAGCGCAACCTGCTCAACCAATCCGTTTACGCCACGACGAGCTTTCT 83285
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
 DB 83284 CTCTGTATACCGCAATGGCGCGAGCTTTTACCGCGCGCTGAACGTGAAGTTT 83230
 RESULT 6
 ID AA253327 standard; DNA; 2277 BP.
 XX
 AC AA253327;

XX 21-MAR-2000 (first entry)
 DT Neisseria meningitidis ORF 149 partial DNA sequence SEQ ID NO:603.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy; ds.
 XX Neisseria meningitidis.
 OS WO9957280-A2.
 PN 11-NOV-1999.
 PD 30-APR-1999; 99WO-US03346.
 PF 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR P-PSDB; AA74565.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 PT Claim 7; Page 422; 1453pp; English.
 PS AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 2277 BP; 600 A; 713 C; 545 G; 419 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 2277
 Score: 3922.00 Matches: 742
 Percent Similarity: 98.68% Conservatives: 7
 Best Local Similarity: 97.76% Mismatches: 9
 Query Match: 97.18% Indels: 2
 DB: 21 Gaps: 0
 US-09-936-377-2 (1-758) x AA253327 (1-2277)
 QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuAsnThrProLeu 20
 DB 1 ATGGCAACAACTACACTCAACCCCATTTGTTTATCAATTCCTTTTATCAACACACCCCTC 60
 QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGly 40
 |||||

Db 61 CTGCCCCAAGCGCATGAACACTGAGCAATCGGTGGATTTGGAAACGGTCAGCGCTCGTCGCG 120
Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db 121 AAAAGCCGTCCGGCGCCACAGTCGCGGCTGTGTGCACATTCAGCCGCTCCGACAAAATC 180
Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db 181 ATCTCCGGCGATACCTTCGCGCAAAAGACCGCTCAACTTGGGCGACGCTTGAAGCGCGTA 240
Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
Db 241 CCGGCGATCCACGCTTCGCAATACCGCGCGCGCGCTGTGCTCCGCTCATTCGCGGTCAA 300
Qy 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
Db 301 ACAGCGAGCGGATTAAGTGTGAACCATCACGGCGAAACAGCGGATATGCGGATTTT 360
Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
Db 361 TCGCCCGATCAGCCCATATATGGTAGATACCGCTTGTTCGCAACAGGTCCAAATCCTCGGC 420
Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValalaasp 160
Db 421 GGGCCGGTTACGCTCTTGTACAGCTCGGGCAATGTGGCGGGCTGGTGCATGTTGCCGAT 480
Qy 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
Db 481 GGCAAAATCCCGCAAAAATGCTGAAACAGCGGTATCGGGGAACTCGGATTCGTTTG 540
Qy 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
Db 541 AGCAGCGGCAATCTGGAAAAATCTACGCTCCGCGCATCAATATCGTTGGGCAAAAAC 600
Qy 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
Db 601 TTTGTATTGCACACGGAAGGCTGTACCGCAATCGGGGATTAACGCGTACCGCGTTAC 660
Qy 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
Db 661 CGCAATCTGAAACGGCTGCCCGACAG - CCACGCCGATTCGCAAAACGGGCGAGCATCGGGCT 719
Qy 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThr - TyrSerAspArgArgAspGlnTyr 260
Db 720 GTCCTTGGGTGGCGAAAAAGGTTTTATCGCGTAGCGTACAGCGACCGTCGCGACCAATA 779
Qy 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280
Db 780 TGGTCTGCTGCCACAGCCACGAATAGATGATTGCCACGCCGACATCATCTGCGCAAAA 839
Qy 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
Db 840 GAGCTTGATTAAACAACGCTATTTACGCTTTATCCGACACCTGTGACCGAAGAAGACAT 899
Qy 300 LasPtyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
Db 900 CGATTAACGACATCCGGGCTTGAGCTGGGCTTCCACGACGAGATATGACACGCGACA 959
Qy 320 sAlaHisAsnGlyLysProTyrPheAspLeuArgAsnLysArgTyrGluLeuArgAlaG 340
Db 960 CACCCACAGCGCGACAGCGTGGATAGACTCGCGCAACAAACGCTACGAACTCCGTGCCGA 1019
Qy 340 uTrrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAspTyr 360
Db 1020 ATGGAAGCAACCGTTCCCGGTTTTGAGCGCTGGCGGTACACCTGAAACCGCAACGACTA 1079
Qy 360 rHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
Db 1080 CCGCCACAGCAAAAAGCGGCGATGCAGTGGGATCTTTTAAACCAACCAACCGCAAAA 1139
Qy 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrPglyValG 400
Db 1140 CGCCCGCATCGAGTTGGCGCCACCAACCCATAGGTCGTCTGAAAGGCGAGCTGGGGGTGCA 1199

Qy 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
Db 1200 ATATTTTACAACAAAATTCAGTGTCTTTATCTGCCATATCCGAAAGCGGTTAAACAACCGAT 1259
Qy 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
Db 1260 GCTGTCTGACAAACAAAGTGCACATTTACAGCTTTTTCGGTGTAGAACAGGCAAACTGGGA 1319
Qy 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db 1320 CAACCTTCAGCTTGAAGGAGCGGTACGCGTGGAAAAACAAAAAGGCTCCATTCAGTACGA 1379
Qy 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyValaHi 480
Db 1380 CAAGCATTTGATTCGGGAAAACTACTACACACCCCTCGCCGAGCTCGGCGCGCA 1439
Qy 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLys 500
Db 1440 CCGCCAAAACCCCGCTCATTCGCACCTTTTCGGGCAACTGGTATTTTCAGCCCAACACAA 1499
Qy 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db 1500 ACTCAGCTTGACCGCTCCCATCAGAAACGCTCGCTCAACGCAAGAGCTGTACGCACA 1559
Qy 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
Db 1560 CGGCAAAACAGCTCGCCACCAACACCTTTGAAAGTCGGCAACAAACACCTCAACAAAGAGCG 1619
Qy 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
Db 1620 TTCCAAACAATATCGAACTCGCGCTGGGTACGAAGCGCACCGCTGCGCAATCAATCTGCG 1679
Qy 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgG 580
Db 1680 ACTTACCGCAACCGCTTCGGTACTACTATTACGCCCAAACTTAAACGACGCGACGCGG 1739
Qy 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db 1740 CCCCATAATCCATCGAAGACGACGAGCAATGAAGCTCGTGGCTACACCAATCCGGCGC 1799
Qy 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG 620
Db 1800 CGACTTCTACGCGCGGGAAGCGAAATCTACTTCAACCGACACCGCTACCGCATCGG 1859
Qy 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgG 640
Db 1860 COTTTCCGGCGCACTATGTACGAGCGCGTCTGAAAAACCTGCTTCCCTACCCCGGAGAGA 1919
Qy 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
Db 1920 AGATSCCTACGCAACCGCTCTTTCATCGCACGACGACCAAAATGCCCCCGTGTTC 1979
Qy 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
Db 1980 GGTCTCGCGCTCGCGCTTCCACCTGAAAGCGCTCGCTGACCGACCGCTATCGATGCAATTT 2039
Qy 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG 700
Db 2040 GGACTACTACCGGTGTTCGCCCAAAACAACTCGCCCGCTACGAAACCGGACCCCGG 2099
Qy 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAs 720
Db 2100 ACACCATATGCTCAACCTCGCGCAAACTACCGCGCAATACCGCTATGGCGAGTGGNA 2159
Qy 720 nTrrTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
Db 2160 TTGGTACGTCAAGCGCAACACCTGCTCAACCAATCCGTTTACGCCCAACAGCAGCTTCT 2219
Qy 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
Db 2220 CTCGTATACGCCCGCAATGGCGCGGAGCTTTTACCGCGCGCGTGAACGTGAAGTTT 2274

RESULT 7	DB:	21	Gaps:	0
AAA81490/c	US-09-936-377-2 (1-758) x AAA81490 (1-1437668)			
ID AAA81490 standard; DNA; 1437668 BP.				
AC AAA81490;				
XX 04-DEC-2000 (first entry)				
XX N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.				
DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;				
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;				
KW Meningococcus B; MenB; ds.				
XX Neisseria meningitidis.				
OS Neisseria meningitidis.				
XX WO200022430-A2.				
FN 20-APR-2000.				
XX 08-OCT-1999; 99WO-US23573.				
XX 09-OCT-1999; 98US-0103794.				
PR 30-APR-1999; 99US-0132068.				
XX (CHIR) CHIRON CORP.				
PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;				
XX Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;				
PI Rappuoli R, Pizza M;				
XX WPI; 2000-318079/27.				
DR Isolated nucleotide sequences of Neisseria meningitidis which can be				
XX used in the diagnosis and treatment of N. meningitidis infection and				
PT other Neisserial infections, for example, N.gonorrhoea -				
XX Claim 7; Page 866-1272; 1760pp; English.				
PS The present invention describes methods of obtaining immunogenic				
XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414				
CC represent specifically claimed Neisseria meningitidis genomic DNA				
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent				
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to				
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the				
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to				
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF				
CC sequences, which are all used in the exemplification of the present				
CC invention. The nucleic acid sequences, protein sequences, and antibodies				
CC against them, can be used in the manufacture of a composition. The				
CC composition can be used as a medicament (or in the manufacture of a				
CC medicament) for treating, preventing or diagnosing infection due to				
CC Neisserial bacteria. For example, some of the identified proteins could				
CC be components of vaccines against Meningococcus B; against all serotypes;				
CC and/or against all pathogenic Neisseriae. Identification of sequences				
CC from the bacterium will also facilitate production of biological probes,				
CC particularly organism-specific probes. Attempts to make efficacious				
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.				
CC Multivalent vaccines have also been tried but none have successfully				
CC overcome antigenic variability. The provision of further, complete				
CC sequences may provide an opportunity to identify secreted or surface				
CC exposed proteins that may be presumed targets for the immune system and				
CC which are not antigenically variable or at least more conserved than				
CC other more variable regions.				
XX Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;				
SQ Alignment Scores:				
Pred. No.:	0	Length:	1437668	
Score:	3922.00	Matches:	742	
Percent Similarity:	98.68%	Conservative:	7	
Best Local Similarity:	97.76%	Mismatches:	9	
Query Match:	97.18%	Indels:	2	

984483 ATGGAGCAACCGTTCCCGGTTTGAAGCCCTCGCGTACACCTGAACCGCAACGACTA 984424
 QY 360 rHisHisAspGluLeuLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
 Db 984423 CGCCACGACGAAAGAGGAGGATGAGTGCAGTTCGAAACTTTTAAACAACCAACGCAAAA 984364
 QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValG1 400
 Db 984363 CGCCCGCATCGAGTTGCGGCACCAACCCCATAGTCTGCTGAAGAGCAGCTGGGGCGTGCA 984304
 QY 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProWe 420
 Db 984303 ATATTTCACACAAATAATCCAGTCTTATTCGCATATCCGAGCGGTAAACACCCGAT 984244
 QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
 Db 984243 GCTGCTTGACACAAAGTGCACATTTACAGCTTTTTCGGTGTAGACAGCAAACTGGGA 984184
 QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
 Db 984183 CAACCTTCACGCTTGAAGAGGCGTACGCTGGGAAAACAAAAGCTTCATTTCAGTACGA 984124
 QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHi 480
 Db 984123 CAAAGCATTTGATTGATCGGGAAACTTACTACAAACCAACCCCTCGCGCACTCGGCGCGCA 984064
 QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisIy 500
 Db 984063 CCGCCAAACCGCCGCTCAATTCGCACTTTTCGGCAACTGGTATTTTCCGCCCAACACACAA 984004
 QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
 Db 984003 ACTCAGCTCAGCCGCTCCCATCAGGAAACGCTCGCGTCAACGCAAGAGCTGTACGACA 983944
 QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisIleAsnLysGluAr 540
 Db 983943 CGGCAACACGCTCGCCACCAACCTTTGAAGTTCGGCAACAAACACTCAACCAAGAGCGG 983884
 QY 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
 Db 983883 TTCCACAAATATCGAATCGCTCGGCTAGCAAGCGCGCTGCGCAATACATCTGCGC 983824
 QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgI 580
 Db 983823 ACTCTACCGAACCGCTTCGCTAACATCATTTACGCCCAACCTTAAACGACGCGCGG 983764
 QY 580 yProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
 Db 983763 CCCCAATCCATCGAAGACGACAGCGAAATGAAGCTGCTGCGCTACACCAATCCGCGCGC 983704
 QY 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG1 620
 Db 983703 CGACTTCTACGCGCGGAAGCGGAAATCTACTTCAACACCGACACCCGCTACCGCATCGG 983644
 QY 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgG1 640
 Db 983643 CGTTTTCGGGACATATGTACAGAGCGCTCTGAAATAACCTTCCTTCCCTACCCGCGAGAG 983584
 QY 640 uAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
 Db 983583 AGATGCTCTACGCGCAACCGTCTCTTTCATCGCACAGGACGACCAAAATGCCCGCGGTTC 983524
 QY 660 oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaSerLe 680
 Db 983523 GCTCGCGCGCTTCGCTTCCACTTGAAAGCTTCCTGACCGACCGTATCGATCGCAATTT 983464
 QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG1 700
 Db 983463 GGACTACTACCGCTGTTTCGCCCAACAAACTCGCCCGCTACGAAACGCGCACGCCCGG 983404
 QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyLutTrpAs 720

Db 983403 ACACCATATGCTCAACCTCGCGCAAACTACCGCCCAATACCGCTATGCGGAGTGGAA 983344
 QY 720 nTyrTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 Db 983343 TTGCTAGTCTCAAGCCGACAACTGCTCAACCAATCCGTTTACGCCCAACAGCAGCTTCT 983284
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 983283 CTCGTATACCCCGCAATAGCGCCGCGACTTACCGCGCGGTGACGCTGAAGTTT 983229
 RESULT 8
 AAZ53328
 ID AAZ53328 standard; DNA; 2277 BP.
 XX
 AC AAZ53328;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 149 partial DNA sequence SEQ ID NO:605.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 03-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR P-PSDB; AAU74566.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 7; Page 424; 1453pp; English.
 XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAU74253 to AAU75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ54673 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 2277 BP; 598 A; 721 C; 547 G; 411 T; 0 other;
 Alignment Scores:

Pred. No.:	0	Length:	2277
Score:	3920.00	Matches:	743
Best Local Similarity:	98.55%	Conservative:	5
Query Match:	97.89%	Mismatches:	10
DB:	21	Indels:	2
		Gaps:	0
US-09-936-377-2 (1-758) x AAZ53328 (1-2277)			
Qy	1	MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu	20
Db	1	ATGGCAAAATACACACCAACCCATTTGTTTATCAATTTCTTTAATCAACACACCCCTC	60
Qy	21	LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly	40
Db	61	CTCTCCCAAGCGCATGGAACTGAGCAATCATGGGCTTGGAAACCGTCAGCGTCGTGGC	120
Qy	41	LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle	60
Db	121	AAAAGCGTCGCGGCGCACTTCGGGCTGCTGCACATTTCTACCGCTCCGCAAAATC	180
Qy	61	IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal	80
Db	181	ATCAGCGCGCACACCTTCGCAAAAGCCGTCACCTTGGGTGATGCTTTAGACGGCGTA	240
Qy	81	ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln	100
Db	241	CCGGGCATTTGCTCGCAATACGGCGCGGCATCCGCTCCGTTATTCGGGGTCAA	300
Qy	101	ThrGlyArgGlnLysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPhe	120
Db	301	ACAGCGAGCGGATTAAAGTGTGAACCATCACGGCGAAACGGGCGACATCGCGGACTTC	360
Qy	121	SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg	140
Db	361	TCTCCAGACCAATGCAATCATGTGTGACAGCGCCCTTGTGCAACAGGTGCAAAATCCTGCGC	420
Qy	141	GlyProValThrLeuLeuTyrSerSerGlyValAsnValAlaGlyLeuValAspValAlaAsp	160
Db	421	GGTCGGTTAGCTCTTGTAAGTCTGGGCAATGTGGCGGGCTGGTGGATTTGCGCGAT	480
Qy	161	GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuLeuGlyLeuArgLeu	180
Db	481	GGCAAAATCCCGAAAAATGCTGAAACCGGCTATCGGCGGAATCGGATTTGGTTTG	540
Qy	181	SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn	200
Db	541	AGCAGCGGCAATCTGGAAAAATCTCAGCTCGGCGGCGCATCAATATCGGTTTGGCAAAAAC	600
Qy	201	PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr	220
Db	601	TTTGATTTGCACACGGAGGCTGTACCGCAATTCGGGGGATTTACGCGTACCGCGTTAC	660
Qy	221	ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla	240
Db	661	CGCAATCTGAAACGCTGCCCGACAG-CCACGCGGATTCGCAACGGGCGACATCGGGCT	719
Qy	241	ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr	260
Db	720	GTCTTGGGTTGGCAAAAAGGCTTTATCGGGCGGACATACAGCGACCGTCGCGACCAATA	779
Qy	260	rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnIly	280
Db	780	TGGTCTGCTGCCACACGACGATAGATGATTTGCCCGCCGACATCATCTGGCAAAA	839
Qy	280	sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa	300
Db	840	GAGTTTGATTAACAAACGCTATTTCAGCTTTATCCGCACTGTTACCGGAAGAAGCAT	899
Qy	300	1AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis	320
Db	900	CGATTACGAATCCGGGCTTGAGCTGGGCTTTTCAGCAGACGATGATGCACACGCCCA	959

Qy	320	sAlaHisAsnGlyLysProIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGln	340
Db	960	TGCCCCAACAGCGCAAACTTGGTAGACTTCGCAACAAACGCTACGAACCTCGCGCGCA	1019
Qy	340	uTriLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr	360
Db	1020	ATGGAAGCAACGTTCCCGGTTTTGAAGCCCTCGCGGTACACCTGAAACCGCAACGACTA	1079
Qy	360	rHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsnGlnThrGlnAs	380
Db	1080	CGCCACGACGAAAAAGCAGCGCATGCACTAGTAGAAAACTTTTAAACAACCAACGCAAA	1139
Qy	380	nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGln	400
Db	1140	CGCCGCTATCGAGTTGGCCACCAACCCATAGCGCTCTGAAAGCGCAGCTGGGGGTGCA	1199
Qy	400	nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe	420
Db	1200	ATATTTGGGACAAAAATCCAGTCTTTATCTGCCACATCCGAAGCGGTCAAAACAACCGAT	1259
Qy	420	tLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAs	440
Db	1260	GCTGCTTGACAAATAAGTGCAACATTACAGCTTTTTCGTTGTAGAACAGGCAAACTGGGA	1319
Qy	440	pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs	460
Db	1320	CAACTTCAGCTTGAAGCGCGCTACGCGTGGAAAAACAAAAAGCCTCATCCGCTACGA	1379
Qy	460	pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHis	480
Db	1380	CAAAAGCATTTGATTTGATCGGGAAAACTACTACCATCCCTGCCCGACCTTCGGCGGCA	1439
Qy	480	sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLys	500
Db	1440	CGGCAAAACCGCCGCTCATTCGCACCTTTTCGGGCAACTGGTATTTTCACGCCACACAA	1499
Qy	500	sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis	520
Db	1500	ACTCAGCCTTGACCGCCCTCCCATCAGAACCGCTCGCTCAACGCAAGAGCTGTACGCACA	1559
Qy	520	sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg	540
Db	1560	CGGCAAAACGCTGCGCAACCAACACTTTGAAGTCGGCAACAAACACCTCAACAAAGAGCG	1619
Qy	540	gSerAsnAsnIleGluLeuAlaLeuGlyTyrGlyAspArgTrpGlnTyrAsnLeuAl	560
Db	1620	TTCCAAACAATATCGAATCTCGCGCTGGCTACGAAGGCGACCGCTGGCAATCACTCTGC	1679
Qy	560	aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGln	580
Db	1680	ACTCTACCGCAACCGCTTCGGCAACTACATTTACGCCCAACCTTAAACGACGACGCGG	1739
Qy	580	yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl	600
Db	1740	CCCCAATTCATCGAAGCAGCAGCAAAATGAGCTGTGCGCTACAAACCAATTCGGTGC	1799
Qy	600	aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGln	620
Db	1800	GGACTTCTACGGCGCGAAGCGGAAATCTACTTCAACCGCACACCGCGCTACCGCATCGG	1859
Qy	620	yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGln	640
Db	1860	CGTTTCGGGCGACTATGTACGAGCGCTCTGAAAAAACCTGCTCTCCCTTACCGCGCAGGA	1919
Qy	640	uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr	660
Db	1920	AGAGCTCTACGGCAACCGCCCATCTCATTTGCCAAGCGCAACCAACGCCCTTCGCGTTC	1979
Qy	660	oAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe	680
Db	1980	GGCTGCGCGCTCGCGCTCCACCTGAAAGCTCGCTGACCGACCGCATCGATGCAATTT	2039
Qy	680	uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGln	700

Db 2040 GGACTACTACCGCGTGTTCGCCCAAAACAACTCGCCCGCTACGAAACGCGCACGCCCG 2099
 QY 700 YHSHLSMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrpAs 720
 Db 2100 ACACCATATGCTCAACTCGCGCAAACTACCGCGCGCAATACCGCTATGGCGAGTGGAA 2159
 QY 720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
 Db 2160 TTGGTACGTCAAAACCGCGCAACCTGCTCAACCAATCCGTTTACGCCCCACAGAGCTTCT 2219
 QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2220 CTCTGATACCGCGCAAAATGGCGCGAGCTTTACCGCGCGCGTGAACGTGAAGTTT 2274

RESULT 9
 ID AAZ53326
 XX AAZ53326 standard; DNA; 2277 BP.
 AC AAZ53326;
 DT 21-MAR-2000 (first entry)
 XX Neisseria gonorrhoeae ORF 149 partial DNA sequence SEQ ID NO:601.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 OS Neisseria gonorrhoeae.
 PN WO9957280-A2.
 XX 11-NOV-1999.
 PD 30-APR-1999; 99WO-US09346.
 PF 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 DR P-P8DB; AAY74564.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX Claim 7; Page 421; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.
 XX Sequence 2277 BP; 599 A; 726 C; 544 G; 408 T; 0 other;

Alignment Scores:
 Pred. No.: 0 Length: 2277
 Score: 3638.00 Matches: 726
 Percent Similarity: 96.84% Conservative: 9
 Best Local Similarity: 95.65% Mismatches: 23
 Query Match: 95.09% Indels: 2
 DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAZ53326 (1-2277)

QY 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuThrProLeu 20
 Db 1 ATGGCCAAATCACACTCAACCCATTTATCAATTCCTTTAATCAACACACCCCTC 60
 QY 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValValGly 40
 Db 61 CTCGCCAAAGCGCATGAAACTGAGCAATCGTGGGCTTGGAAACGTCAGCGTCGTCGC 120
 QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 121 AAAAGCGTCCGCGCGGACTTCGGGGCTGTGCACACTTCGACCGCTCCGCAAAATC 180
 QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 Db 181 ATCTCGGCGATACITTTGGCCCAAAAGCCGTCACTTGGCGACCTTTGGACGGCGTA 240
 QY 81 ProGlyIleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 241 CCGGGCATCCACGCTTCGCAATACGGCGCGCGGCGATCCGCTCCCTTATTTCGGCTCAA 300
 QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
 Db 301 ACGGCGAGACGGATTAAGTATTGAACCATCACGGGAAACGGCGGATATGGCGGACTTT 360
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 Db 361 TCTCCGATCAGCCCATATTAGTATACCGCTTGTCCAAACAGGTTTGAATTCCTCGCG 420
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
 Db 421 GGCGCGGTACGCTCTTGTACAGCTCGGCAATGTGGCGGGCTGTGTTCGATGTTCGGAT 480
 QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeu 180
 Db 481 GGAATAATCCCGAAAAATGCTGAAACGGCGTATCGGGCGGAAACCGGATTCGCTTTG 540
 QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
 Db 541 AGCAGCGCAATTTAGAAAACTGACATCCGAGGATCAATATCGGACTGGGCAAAAAC 600
 QY 201 PheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
 Db 601 TTCGTGCTCATACCGAAGGCTTTGTACGCAATCGGGCGATTACGCCGCTACCGCGTTAC 660
 QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
 Db 661 CCAATCTGAACCGCTGCCGACAG-CCATCCCGATTCGCAACCGGCGAGCATCGGGCT 719
 QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThr-TyrSerAspArgArgAspGlnTyr 260
 Db 720 GTCTTGGTGGGCGAAAAAGGCTTTATCGGCGACATACAGCGACCGTCCGCGCGCTA 779
 QY 260 rGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnly 280
 Db 780 CGGCGCTGCTGCCACAGCCAGCAATACGATGATTGCCAGCGCGCATCATCTGGCAAAA 839
 QY 280 sSerLeuLeuHisArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVa 300
 Db 840 GAGTTTGATCAACAAACCGCTATTTCGAGCTTTATCCGCACTTGTGTGACCGAAGACAT 899

```

QY 300 lAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
Db :|||||
QY 900 CGATTACGACAAATCCGGGCTTGAGCTGGCGCTTCACGACGGCGCGGTGCGACACGCGACA 959
Db :|||||
QY 320 sAlaHisAsnGlyLysProTTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaG1 340
Db :|||||
QY 960 CACCCACAACGGCAACCGTGATAGACTCGCGCAACAAACGCTACGAACTCCGGCGCGCA 1019
Db :|||||
QY 340 uTrpLysGlnPropheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTy 360
Db :|||||
QY 1020 ATGGAAGCAGCGCATTCGCCGGTTCGAGCCCTGGCGGTACATCTGAACCGCAATGACTA 1079
Db :|||||
QY 360 rHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAs 380
Db :|||||
QY 1080 CCACCACGACGAAAGCAGCGGATGCGAGTGAAGAACTTCCTTCAACAAACAAACACACAA 1139
Db :|||||
QY 380 nAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValG1 400
Db :|||||
QY 1140 CGCCCGTATCGAGTTGGCCACCAACCCCATAGCCGCTCTGAAGGCGAGCTGGGGCGTGCA 1199
Db :|||||
QY 400 nTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMe 420
Db :|||||
QY 1200 ATATTGGGACAAAATCCAGCGCGCTTTCGCCCATTTCCCGAAACCGTCCCAACACCGAT 1259
Db :|||||
QY 420 tLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAs 440
Db :|||||
QY 1260 GTTGATTGACAAACAATGTCGCCATTTACGCTTTTCGGTGAGAACAGCGAAATTTGGGA 1319
Db :|||||
QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db :|||||
QY 1320 CAACITTCAGCTTGAAGCGCGGTACGCGTGAAACAAACAAAGCCTCCATCCGCTACGA 1379
Db :|||||
QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHis 480
Db :|||||
QY 1380 CAAGCATTTGATGATCGAGAAACATCTACAAACAGCGCCCTCGCGCGCGCA 1439
Db :|||||
QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLys 500
Db :|||||
QY 1440 CGCCCAACCGCCGCTGCTGTCGACTTCGGGCAACTGGTATTTACGCCACACACACA 1499
Db :|||||
QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis 520
Db :|||||
QY 1500 ACTCAGCTGACCGGCTCCCATCAGAACCGCTCGCTCAACGCAAGAACTGTACGCGACA 1559
Db :|||||
QY 520 sGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluAr 540
Db :|||||
QY 1560 CGGCAAGCAGCGTCGCCACCAACACCTTTGAAGTCGCGCAACAAACACCTCAACAAAGAGCG 1619
Db :|||||
QY 540 gSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAl 560
Db :|||||
QY 1620 TTCCAAACATATCGAACTCGCGCTGGGTACGAAGCGACCGCTGGCAATACATCTGCG 1679
Db :|||||
QY 560 aLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgG1 580
Db :|||||
QY 1680 AGCCTACGCGCAACCGATTCGGCACTACATTTACGCCCAACCTTAAACGACGCGACGCG 1739
Db :|||||
QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db :|||||
QY 1740 CCCCAATTCATCGAAGACGACGCGAATGAAGCTCGTGGCTACAAACCAATCCGGTGC 1799
Db :|||||
QY 600 aAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleG1 620
Db :|||||
QY 1800 CGACTTCTACGGCGCGGAAGCGGAATCTACTTCAACCGACACCGCTACCGCATCGG 1859
Db :|||||
QY 620 yValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgG1 640
Db :|||||
QY 1860 TGTTTTCCGGCGCACTATGTACGAGGCGGTCTGAAAAACCTCGCGTCCCTACCGCGGAGGA 1919
Db :|||||
QY 640 uAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPr 660
Db :|||||
QY 1920 AGATCCCTACGGCAAAACGTCCCTTCATCGCAACGCCGACCAAAACGCGCCCGCATTC 1979

```

```

QY 660 ohlaalaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLe 680
Db :|||||
QY 1980 GGCTCGCGGCTCGGCTTCCACCTGAAAACCTCGTAAACCGGCTATCGATGCCAATTT 2039
Db :|||||
QY 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG1 700
Db :|||||
QY 2040 GGACTACTACCGGTGTTCCGCCAAACAACTCGCCGCTACGAACCGGTACGCCCGG 2099
Db :|||||
QY 700 yHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyLutPAs 720
Db :|||||
QY 2100 ACACCATATGCTCAACCTCGGTGCAAACTACCGCGCAATACGCGCTATGGCGAGTGAA 2159
Db :|||||
QY 720 nTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLe 740
Db :|||||
QY 2160 TTGGTACGTCAAGCGCAACCTGCTCAACCAATCCGTTTACGCCCCACAGCAGCTTCT 2219
Db :|||||
QY 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db :|||||
QY 2220 CTCTGATACCGCGCAATGGGCGCGAGCTTTACCGCGCGGTAAACGTGAAGTTT 2274
Db :|||||
RESULT 10
AAZ53318
ID AAZ53318 standard; DNA; 2208 BP.
XX
AC AAZ53318;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 147 partial DNA sequence SEQ ID NO:585.
XX
KW Neisseria meningitidis; Neisseria gonorrhea; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
XX
PR 31-JUL-1998; 98US-0094869.
XX
PR 02-SEP-1998; 98US-0098994.
XX
PR 02-SEP-1998; 98US-0099062.
XX
PR 09-OCT-1998; 98US-0103749.
XX
PR 09-OCT-1998; 98US-0103794.
XX
PR 09-OCT-1998; 98US-0103796.
XX
PR 25-FEB-1999; 99US-0121528.
XX
(CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
PA
Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; AAY74556.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
PT
PS Claim 7; Page 412-413; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhea polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the

```

CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX
SQ Sequence 2208 BP; 579 A; 689 C; 540 G; 400 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 2208
Score: 3812.00 Matches: 718
Percent Similarity: 98.78% Conservative: 8
Best Local Similarity: 97.69% Mismatches: 8
Query Match: 94.45% Indels: 2
DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAZ53318 (1-2208)

QY 25 HisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
DB 4 CATAAACTGAGCAATCGGTGGATTTGGAAACGGTCAGCGTCGTGGCAAAAGCGGTCCG 63
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
DB 64 CGGCCACGTCGGGGCTGTTGCACACTTCGACCGCTCCGACAAATCATCTCCGGCGAT 123
QY 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
DB 124 ACCTTGGCCAAAAGCGGTCACTTGGGGACGCTTACAGCGCGTACCGGGCATCCAC 183
QY 85 AlaSerGlnTyrglyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
DB 184 GCTTCGCAATACCGCGCGCGCGCTGTGCTCCGCTCATTCGCGGTCAAAACAGGCGCGG 243
QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
DB 244 ATTAAAGTGTGAACCATACCGCGGAACAGGGCGATATGCGGATTTTCGCGCGATCAC 303
QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
DB 304 GCCATTATGTAGATACCGCTTGTGCAACAGGTGCAAAATCTCGCGGGCGGTACG 363
QY 145 LeuLeuTyrsSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
DB 364 CTCCTGTACAGCTCGGGCAATGTGGCGGGCTGTGCTGATGTTGCCGATAAATCCCC 423
QY 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
DB 424 GAAAAATGCCTGAAACCGCGTATCGGGCAACTCGGATGCGTTTGACGCGGCAAT 483
QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
DB 484 CTGGAAAACTCACGTCCGGCGCATCAATATCGGTGGGCAAAATTTGTATTGCAC 543
QY 205 ThrGluGlyLeuTyrglySerGlyAspTyraValProArgTyrglyArgAsnLeuLys 224
DB 544 ACGGAAGGGCTGTACCGCAAAATCGGGGATTTACGCGGTACCGGTATCTGAA 603
QY 225 ArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTTP 244
DB 604 CGCCTGCCCGACAG-CCACGCCGATTCGCAACGGCGCAGCATCGGGCTGTCTTGGGTGG 662
QY 245 ArgLysArgPheTyraArgThr-TyrSerAspArgAspGlnTyrglyLeuProAl 264
DB 663 CGAAAAAGGTTTTATCGGGTAGCGTACAGCGACCGCTCGCGACCAATATGTTGCTGCTGC 722
QY 264 aHisSerHisGluTyraAspCysHisAlaAspIleIleTTPGlnLysSerIleuLeAs 284
DB 723 CCACGCCACGAAATACGATGATTGCCACCGCGACATCATCTGGCAAAAGAGCTTGATTAA 782
QY 284 nLysArgTyraLeuGlnLeuTyraProHisLeuLeuThrGluGluAspValAspTyraSpAs 304

DB 783 CAAACGCTATTATTACAGCTTTATCCGCACTTGTGGACCGAAGACATCGATTACGACAA 842
QY 304 nProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGly 324
DB 843 TCCGGGGCTTGAAGTGGGGCTTCCAGCAGCAGTAATATGTCACGACACACCCACACGG 902
QY 324 YLysProTTPileAspLeuArgAsnLysAsTyrglyLeuArgAlaGluTyraLysGlnPr 344
DB 903 CAGACCGTGATAGACTCGGCAACAAACGCTACGAACTCGTGCCGATATGGAAGCAACC 962
QY 344 oPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyraHisAspGly 364
DB 963 GTTCCCGGGTTTGAAGCCCTGGCGGTACACTTGAACCGCAACGACTTACCGCCACGACGA 1022
QY 364 uLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGly 384
DB 1023 AAAAGCAGGCGATGCAGTCGAAACCTTTTAAACCAACCAACCAACCAACCGCATCGA 1082
QY 384 uLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrglyValGlnTyraLeuGlyGly 404
DB 1083 GTTGGCCCAACCACTAGTCTGTGAAGAGCAGCTGGGGGTGCAATATTTTACAAACA 1142
QY 404 nLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
DB 1143 AAAATCCAGTCTTTATCTGCCATATCCGAAGCGGTAAACCAACCGATGCTGCTTGACAA 1202
QY 424 nLysValGlnHisTyraSerPheGlyValGluGlnAlaAsnTyraAspAsnPhethrLe 444
DB 1203 CAAAGTGCACATTTACGCTTTTTCGGTGTAGAACAGGCAAACTGGGACAACTTCAGCT 1262
QY 444 uGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyraAspLysAlaLeuI 464
DB 1263 TGAAGGAGGCGTACGCGTGGAAAAACAACAAAGGCTCCATTCAGTACGACAAAGCATTGAT 1322
QY 464 eAspArgGluAsnTyraTyraLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAl 484
DB 1323 TGAATCGGAAAACTACTACAAACCCCTGCGCGACCTCGCGCGCGCACCGCAACCGC 1382
QY 484 argSerPheAlaLeuSerGlyAsnTyraPheThrProGlnHisLysLeuSerLeuth 504
DB 1383 CCGCTCATTCGCACTTTCGGGCAACTTGTATTTTCAGCCCAACACACAACTCAGCGCTGAC 1442
QY 504 rAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyraAlaHisGlyLysHisVa 524
DB 1443 CGCCTCCCATCAGGAACGCTCGCGTCAACGCAAGAGCTGTACGACACGCGCAACACGT 1502
QY 524 lAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnI 544
DB 1503 CCGCCACCAACACCTTTGAAAGTCGGCAACAAACACCTCAACAAAGAGCGTTCCCAACAATAT 1562
QY 544 eGluLeuAlaLeuGlyTyraGlyAspArgTTPGlnTyraAsnLeuAlaLeuTyraArgAs 564
DB 1563 CGAATCTGGCGTGGGTACGAGCGCGACCGCTGGCAATCAATCTGGCACTCTTACCGCAA 1622
QY 564 nArgPheGlyAsnTyraIleTyraAlaGlnThrLeuAsnAspGlyArgGlyProLysSeril 584
DB 1623 CCGCTTCGGTAACTACATTTACGCCCAACCTTAAACGACGCGCGCGCCCAAAATCCAT 1682
QY 584 eGluAspAspSerGluMetLysLeuValArgTyraAsnGlnSerGlyAlaAspPheTyraGly 604
DB 1683 CGAAGACGACAGCAAAATGAAGCTGTGCGCTACAAACCAATCCGCGCGCGACTTTCACGG 1742
QY 604 yAlaGluGlyGluIleTyraPheLysProThrProArgTyraArgIleGlyValSerGlyAs 624
DB 1743 CGCGGAAGCGGAATCTACTTCAAACCGACACCGCGCTACCGCATCGGCGTTTCGCGCGA 1802
QY 624 pTyraValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyraGly 644
DB 1803 CTATGTAGGAGCGCGTCTGAAACAACTGCTTCCCTCCCTCCCGCAGAGAGATGCCCTACGG 1862
QY 644 yAsnArgProPheIleAlaGlnAspGlnAsnAlaProArgValProAlaAlaArgLe 664
DB 1863 CAACCGTCTTTTCATCGCAGCAGGACGCAAAATGCCCCCGCTGTTCCGCGCTGCGCGCT 1922

QY 264 aHisSerHisGluTyrAspAspCysHisAlaAspIleTleTTrpGlnLysSerLeuIleAs 284
 DB 22141 CCACAGCCACGAATACATGATGTCACCGCCGACATCATCTGGCAAAAGAGCTTGATTAA 22082
 QY 284 nLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAs 304
 DB 22081 CAACCGCTATTACAGCTTTATCCGACCTGTTGACCGAAGAAGACATCATGATTACGACA 22022
 QY 304 nProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAlaHisAsnGln 324
 DB 22021 TCCGGGCTTGAGCTGGGCTTCACAGCAGCATGATGACACGACACACACACGAGCGG 21962
 QY 324 yLysProTropIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnPr 344
 DB 21961 CAGACGCTGGATAGACTGGGCAACAAACGCTACGAACTCGGTGGCATGGAGCAAC 21902
 QY 344 opheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGln 364
 DB 21901 GTTCCCGGCTTTGAAGCCCTGGCGGTACACTGAACCGCAACGACTACGCGCACGACA 21842
 QY 364 uLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGln 384
 DB 21841 AAAGCAGGCGATGCGATCGAATCGAAACCTTTTAAACCAACCAACGCAACGCGCATCGA 21782
 QY 384 uLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGln 404
 DB 21781 GTTGGCGCCACCAACCCATAGGTGCTCTGAAGGCGAGCTGGGGCGTGCAATATTACACA 21722
 QY 404 nLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
 DB 21721 AAATCCAGTGCTTTATCTGCCATATCCGAAGCGGTTAAACCAACCGATGCTGCTGACAA 21662
 QY 424 nLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLe 444
 DB 21661 CAAGTGCACATATACAGCTTTTTCGGGTAGAACAGGCAAACTGGGCAACACTTCACGCT 21602
 QY 444 uGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle 464
 DB 21601 TGAAGGAGGCGTACGCGTGGAACAAAGAGCTTCATTCAGTACGACAAACATTTGAT 21542
 QY 464 eAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyValAlaHisArgGlnThrAl 484
 DB 21541 TGATCGGAAACACTACTACACACACCCCTGCGGACCTCGCGCGCACCGCCAAACCGC 21482
 QY 484 aArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuTh 504
 DB 21481 CCGCTCATTCGCACTTTGCGGCAACTGGTATTTACGCCCAACACACAACTCAGCCTGAC 21422
 QY 504 rAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisVa 524
 DB 21421 CGCTCCCATCAGAACCGCTGCGCTCAACGCAAGAGCTGTACGACACGCGCAACACGT 21362
 QY 524 lAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIle 544
 DB 21361 CGCCACCAACACCTTTGAAGTCGCGCAACAAACACTCAACAAAGAGGCTTCCCAACAATAT 21302
 QY 544 eGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAs 564
 DB 21301 CGAACTCGGCTGGGTACGAGCGGACCGCTGCAATACATCTGCACTCTACCGCAA 21242
 QY 564 nArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIle 584
 DB 21241 CCGCTTCGTTAACTACATTTACGCGCCAAACCTTAAACGAGGAGCGCGCCCAAAATCCAT 21182
 QY 584 eGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGln 604
 DB 21181 CGAAGCAGCAGCGAAATGAAGCTGTCGCTACACCAATCCGCGCCGACACTCTACGG 21122
 QY 604 yAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAs 624
 DB 21121 CGCGAAGCGGAAATCTACTTCAAAACGACACCGCGCTACCGCATCGGCGTTCGCGCGA 21062
 QY 624 pTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGln 644

DB 21061 CTATGTACGAGCGCTCTGAAAAACCTGCCTTCCCTACCCGCGAGAGATGCCTACGG 21002
 QY 644 yAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLe 664
 DB 21001 CAACCGCTCTTTCATCGCACAGGACGACCAAAATGCCCGCCCGCTGTCGCGCCT 20942
 QY 664 uGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrAr 684
 DB 20941 CGGCTTCACCTGAAAGCCTCGCTGACCGACCGTATGATGCCAATTTGGACTACTACCG 20882
 QY 684 gValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLe 704
 DB 20881 CGTGTTCGCCCCAAACAACTCGCCCGCTACGAAACGCGACGCGCCGACACCATATGCT 20822
 QY 704 uAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTyrAsnTrpTyrVally 724
 DB 20821 CAACCTCGCGCAACACTACCGCCGCAATACCGCTATGCGGAGTGAATTTGGTACGTCAA 20762
 QY 724 sAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPr 744
 DB 20761 AGCGACAACTGCTCAACCAATCCGTTTACGCCACACGACGCTTCTCTGTATAGGCC 20702
 QY 744 oGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 DB 20701 GCMAATGGCGCGACGCTTTACCGCGCGCGTGAACGTGAAGTTT 20659

RESULT 12

ABZ41893
 ID ABZ41893 standard; DNA; 2184 BP.
 XX AC ABZ41893;
 XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae nucleotide sequence SEQ ID 8375.
 XX DX Antibacterial; infection; vaccine; gene therapy; gene; ds.
 XX KW Neisseria gonorrhoeae.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.
 XX PF 12-FEB-2002; 2002WO-IB02069.
 XX PR 12-FEB-2001; 2001GB-0003424.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Fontana MR, Pizza M, Masignani V, Monaci E;
 XX DR WPI; 2003-058415/05.
 XX DR P-PSDB; ABP80923.
 XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX PT medicament for treating or preventing N. gonorrhoeae infection
 XX PS Disclosure; Page 798-799; 815pp; English.
 XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
 XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
 XX CC antibodies that specifically bind to the proteins. The composition
 XX CC comprising the protein, nucleic acid or antibody is useful for the
 XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 XX CC infection, this may be in the form of a vaccine or gene therapy.
 XX CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
 XX CC molecules of the invention.
 XX SQ Sequence 2184 BP; 568 A; 696 C; 535 G; 385 T; 0 other;
 Alignment Scores:

Pred. No.: 0 Length: 2184
 Score: 3706.00 Matches: 698
 Percent Similarity: 96.98% Conservative: 9
 Best Local Similarity: 95.75% Mismatches: 21
 Query Match: 91.82% Indels: 2
 DB: 25 Gaps: 0

US-09-936-377-2 (1-758) x AB241893 (1-2184)

QY	31	ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu	50
DB	1	GTGGGCTTGAAACCGTCAGCGTCGTGCGCAAAAGCGGTCGCGCGAGCTTCGGGGCTG	60
QY	51	LeuHisThrSerThrAlaSerAspLysIlelleSerGlyAspThrLeuArgGlnLysAla	70
DB	61	CTGCACTTCACCGCGCTCCGACAAATCATCTCCGGCGATCTTTGGCGCAAAAGCC	120
QY	71	ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGly	90
DB	121	GTCACTTCGGCGAGCGCTTTGACGGCGTACCGGGCATCCAGCTTCGCAATACGGCGG	180
QY	91	GlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHis	110
DB	181	GGCGCATCCGCTCCCGTTATTCCGGTCAACCGGCAGACGGATTAAAGTATTGAACCAT	240
QY	111	HisGlyGluThrGlyAspMetAlaaspPheSerProAspHisAlaIleMetValaspThr	130
DB	241	CACGGCGAAACGGCGGATATGGCGACTTTTCTCCGATCACGCCATTATGTAGTATACC	300
QY	131	AlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGly	150
DB	301	GCCTTGTCACAGGTGGAATCTTCGCGGGCGGTTACGCTCTGTACAGCTCGGGC	360
QY	151	AsnValAlaGlyLeuValaspValAlaAspGlyLysIleProGluLysMetProGluAsn	170
DB	361	AATGTGGCGGGCTGCTCGATGTGCGGATGGAAAAATCCCGGAAAAATGCTGAAAC	420
QY	171	GlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlnLysLeuThrSer	190
DB	421	GGCGTATCGGGCGAGCGCGATTGGCTTTGACGCGCGCAATTTGAAAAAATGACATCC	480
QY	191	GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArg	210
DB	481	GCAGGCATCAATATCGACTGGGCAAAACTTCGTGCTGCATACCGAAGCTTTGACCGC	540
QY	211	LysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerPro	230
DB	541	AAATCGGGCGATTACGCGGTACCGGTTACCGCAATCTGAAACGCGCTGCCGACAG-CCA	599
QY	231	ArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArg	250
DB	600	TGCCGATTTCGCAACCGGCGAGCATCGGCTGCTTGGGTGGCGGCAAAAGGCTTTATCGG	659
QY	251	ArgThr-TyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAs	270
DB	660	CGCAGCATACAGCACCGGTGCGACCGCTACGCGCTGCTGCCACAGCCACGAATACGA	719
QY	270	paspcyHisAlaAspIlelleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLe	290
DB	720	TGATTGCCACGCGCATCATCTGGCAAAAGAGTTTGATCAACAAACGCTATTTCAGCT	779
QY	290	uTyrProHisLeuLeuThrGluGluaspValAspTyrAspAsnProGlyLeuSerCysG	310
DB	780	TTATCCGCACTTGTATCCGAAGAAGACATCGATTACGACAAATCCGGGCTTGAGCTGGG	839
QY	310	yPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAspLe	330
DB	840	CTTCCAGCAGCGCGAGCGTGCACACGACACACCCACCAACGGCAAAACCGTGGATAGACT	899
QY	330	uArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAl	350
DB	900	GGCAACAAACGCTACGAACTCCGCGCGAATGGAAGACGCAATCCCGGTTTTTGAAGC	959

QY	350	aLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGlyLeuAlaGlyAspAlaVa	370
DB	960	CCTGGCGGTACATCTGAACCGCAATGACTACCAACGACGCAAAAGCAGCGATGACGT	1019
QY	370	lGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle	390
DB	1020	AGAAACACTTCTTCAACAAACAAACACACAAACCGCGTATCGAGTTGCGCCACCAACCCAT	1079
QY	390	eGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSe	410
DB	1080	AGCCCGTCTGAAGGCGAGCTGGGGCTGCAATATTGGGACAAATATCAGCGCGCTTTC	1139
QY	410	rAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSe	430
DB	1140	CGCCATTCCGAAACCGTCCAAACACCGGATGTGATTGACAAACATGTCCGCCATTACAG	1199
QY	430	rPhePheGlyValGluGlnAlaAsnTyrAspAsnPheThrLeuGluGlyGlyValArgVa	450
DB	1200	CTTTTTCGGTGTAGAACAGGCAAAATGGGACAACTTCACGCTTGAAGGCGGCGTACGCGT	1259
QY	450	lGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTy	470
DB	1260	GGAAAAACAAAGAGCTCCATCCGTCACGACAAAGCATTTGATTGATCGAGAAACACTACTA	1319
QY	470	rLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSe	490
DB	1320	CAACGAGCCCTGCGCGAGCTCGGGCGGCACGCGCAACCGCGCTCGTTCGCACTTTC	1379
QY	490	RGlyAsnTyrTyrPheThrProGlnHisLysLysLeuSerLeuThrAlaSerHisGlnGluAr	510
DB	1380	GGGCAACTGGTATTTCACGCCACACCAAACTCAGCCTGAGCGGCTCCCATCAGGAACG	1439
QY	510	gLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGl	530
DB	1440	CCTGCGCTCAACGCAAGAACTGTACGCACACGGCAAGCACTCGCCACCAACACCTTGA	1499
QY	530	uValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTy	550
DB	1500	AGTCGGCAACAAACACCTCAACAAAGAGCGTTCCAAACATATCGAACTCGCGCTGGGCTA	1559
QY	550	rGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIl	570
DB	1560	CGAAGCGCACCGCTGGCAATACAACTCGGCACGCTACCGCAACCGATTTCGCAACTACAT	1619
QY	570	eTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMe	590
DB	1620	TTACGCCCAAACTTAAACGACGAGCGCGGCCCAAAATCCATCGAAGACGACAGCGAAAT	1679
QY	590	tlLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTy	610
DB	1680	GAAAGCTCGTGGCTACAACCAATCCGTCGCGACTTCTACGCGCGGGAAGCGAAATCTA	1739
QY	610	rPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLe	630
DB	1740	CTTCAAAACGACCGCGCTACCGCATCGGTGTTCGCGGCACTATGTACGAGGCGGCTT	1799
QY	630	uLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAl	650
DB	1800	GAAAAACCTGCGCTCCCTACCGCGGAGGAAAGATCCCTACGCGCAACGCTCCCTTCAICG	1859
QY	650	gGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAl	670
DB	1860	ACAAGCCGACCAAAACGCGCGCATTCGCGCTCGCGCTCGCGCTTCCACCTGAAAC	1919
QY	670	aSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLy	690
DB	1920	CTCGTAAACCGACCGTATCGATGCCAATTTGAGTACTACCGCGTGTTCGCGCAAAACAA	1979
QY	690	sLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyValAsnTy	710
DB	1980	ACTCGCGCGCTACGAAACGCGTACCGCGGACCAATATGTCTCAACCTCGGTGCAAACTA	2039
QY	710	rArgArgAsnThrArgTyrGlyGluTyrTrpAsnTrpTyrValLysAlaAspAsnLeuAs	730

```

Db 2040 CCGCGCAATACGGCGTATGCGAGTGGAAATGGTACGTCAAGCCGACCACTGCTCAA 2099
QY 730 nGlnSerValTyrAlaHisSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 2100 CCAATCGTTTACGCCACAGACGCTTCTCTGTATACGCCGCAAAATGGCGCAGCTT 2159
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 2160 TACCGCGCGGTAAACGTAAGTTT 2184

RESULT 13
AAZ53324
ID AAZ53324 standard; DNA; 1020 BP.
XX
XX AAZ53324;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 149 partial DNA sequence SEQ ID NO:597.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
XX Neisseria meningitidis.
OS
XX
PN WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094859.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 03-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
DR P-PSDB; AAY74562.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX Claim 7; Page 418-419; 1453pp; English.
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 1020 BP; 274 A; 344 C; 226 G; 176 T; 0 other;
SQ

```

```

Alignment Scores:
Pred. No.: 7,94e-153 Length: 1020
Score: 1808.00 Matches: 336
Percent Similarity: 99.41% Conservative: 1
Best Local Similarity: 99.12% Mismatches: 2
Query Match: 44.80% Indels: 0
DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAZ53324 (1-1020)

QY 420 MetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp 439
Db 1 ATGCTGCTTGACAAACAAGTGCACATTTACAGCTTTTTCGGTGTAGAACAGCAAACTGG 60
QY 440 AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyr 459
Db 61 GACAACTTCAGCTTGAGAGGCGGTACGCGTGGAANAACAAAGAGCTCAITTCAGTAC 120
QY 460 AspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyVala 479
Db 121 GACAAAGCATTTGATTCGGGAAACTACTACAACCCACCCCTGCCCGACCTCGGCGCG 180
QY 480 HisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHis 499
Db 181 CACCGCAAAACCGCCGCTCATTCGCACCTTCGGGCAACTGGTATTTTCAGCCCAACAC 240
QY 500 LysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAla 519
Db 241 AAATCTAGCCTGACCGCTCCCATCAGGAACGCGTCCGCTCAAGCAAGAGCTGTACGCA 300
QY 520 HisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGlu 539
Db 301 CACGGCAAAACACGCTCGCCACCAACCTTTGAAGTGGCAACAACAACCTCAACAAAGAG 360
QY 540 ArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeu 559
Db 361 CGTTCCAAACAATATCGAACTCGCGCTGGGCTACGAAAGCGGACCGCTGGCAATCAATCTG 420
QY 560 AlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArg 579
Db 421 GCACCTTACCGCAACCGCTTCGGTACTACATTTACGCCCAACCTTTAAACGACGACGC 480
QY 580 GlyProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
Db 481 GGCCCCAAATCCATCGAAGACGACAGCAAAATGAAGCTCGTGCCTACCAACCAATCCGCGC 540
QY 600 AlaAspPheTyrGlyValaGluGlyGluIleTyrPheLysProThrProArgTyrArgIle 619
Db 541 GCGGACTTCTACGGCGCGGAAGCGAAATCTACTTCAAAACCGACACCGCGCTACCGGATC 600
QY 620 GlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArg 639
Db 601 GCGGTTTCGGCGGACTATGTACGAGGCGCTCTGAAAAACCTGCTTCCCTACCGCGCAGA 660
QY 640 GluAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
Db 661 GAAGATGCTCTACGCGCAACCGTCTCTCATCGCAGACGACCAAAATGCCCCCGGTGT 720
QY 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
Db 721 CCGGCTGCGGCGCTTCGGCTTCACCTGAAAGCTCGCTGACCGCGATTCGATGCGCAAT 780
QY 680 LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
Db 781 TTGGACTACTACGCGGTGTTCGCCCAAAACAACAACTGCCCGCTACGAAACGCGCACGCC 840
QY 700 GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrp 719
Db 841 GGAACCATATGTCTCAACTCGGCGCAAACTACCGCGCAATACGCGCTATGCGGAGTGG 900
QY 720 AsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPhe 739
Db 901 AATTGTGTACGTCAAGCGCCGACCACTTGCTCAACCAATCGGTTTACGCCCCACGACGCTTT 960

```


Qy 740 LeuSerAspThrProGlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
 Db 961 CTCTCTGATACGCCGCAAAATGGCGCGAGCTTTACCGCGCGGTGACGTAAGTTT 1017

RESULT 14

AAZ53325
 ID AAZ53325 standard; DNA; 1020 BP.
 AC
 XX AAZ53325;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 XX Neisseria meningitidis ORF 149 partial DNA sequence SEQ ID NO:599.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX WO9957280-A2.
 FN
 XX PD 11-NOV-1999.
 XX
 XX 30-APR-1999; 99WO-US09345.
 PF
 XX
 XX 01-MAY-1998; 98US-0083758.
 PR
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 09-OCT-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX

(CHIR) CHIRON CORP.

(GENO-) INST GENOMIC RES.

PA Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;

XX WPI: 2000-062150/05.

DR P-PSDB; AAY74563.

XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -

XX Claim 7; Page 420; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX SQ Sequence 1020 BP; 271 A; 351 C; 228 G; 170 T; 0 other;

Alignment Scores:

Pred. No.:	2,67e-151	Length:	1020
Score:	1791.00	Matches:	334
Percent Similarity:	98.53%	Conservative:	0
Best Local Similarity:	98.53%	Mismatches:	5
Query Match:	44.38%	Indels:	0

DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAZ53325 (1-1020)

Qy 420 MetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlnAlaAsnTrp 439
 Db 1 ATGCTGCTTGACAAATPAAAGTGCACATTACAGCTTTTTCGGTGAGAACAGCAAACTGG 60
 Qy 440 AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyr 459
 Db 61 GACAACTTCACGCTTGAAGGCGCGTACGCGTGGAAAAACAAACAAAGCCCTCCATCCGCTAC 120
 Qy 460 AspLysAlaLeuLleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAla 479
 Db 121 GACAAAGCATTTGATTGATCGGGAACACTACTACAACCATCCCTCCGACCTCGCGCG 180
 Qy 480 HisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHis 499
 Db 181 CACGCCAAACCGCCCGCTCATTCGCACCTTTCGGGCAACTGGTATTTACAGCCCAACAC 240
 Qy 500 LysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAla 519
 Db 241 AAACCTCAGCCTGACCGCCTCCCATCAGGAACGCGTCAACGCAAGAGCTGTACGCA 300
 Qy 520 HisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGlu 539
 Db 301 CACGCAAAACACGTCGCCCAACACCTTTGAAGTCGGCAACAAACACACCTCAACAAAGAG 360
 Qy 540 ArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeu 559
 Db 361 CGTTCCAAACATATCGAACTCGCGCTGGCTACGAAAGGCGACCGCTGGCAATACAAATCTG 420
 Qy 560 AlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArg 579
 Db 421 GCACCTCTACCGCAACCGCTTCGGCAACTACATTTACGCCCAACCTTAAACGACGAGCGC 480
 Qy 580 GlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
 Db 481 GGCCCCAAATCCATCGAAGACGACAGCAAGAAATGAAGCTCGTCGCTACAAACCAATCCG 540
 Qy 600 AlaAspPheTyrGlyAlaGluGlyLulleTyrPheLysProThrProArgTyrArgIle 619
 Db 541 CGGACTTCTAGCGCGCGAAGGCGAAATCTACTTCAAACCGCACCGCGCTACCGCATC 600
 Qy 620 GlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArg 639
 Db 601 GCGGTTTCCGCGCACTATGTACGAGCGCTCTGAAAAACCTGCCTTCCCTACCCGCGAG 660
 Qy 640 GluAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
 Db 661 GAAGACGCTTACGGCAACCGCCCACTCATTCGCCAAGCGCACCAAAACGCGCTCGCGTT 720
 Qy 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
 Db 721 CCGGCTGGCGGCTTCGGGCTCCACTGAAGCCTCGCTGACCGCACCGCATCGCATCCCAAT 780
 Qy 680 LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
 Db 781 TTGACTACTACCGCGTGTTCGCCCAAAACAACTCGCCCGCTACGAAACGCGCACGCC 840
 Qy 700 GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrp 719
 Db 841 GGACACCAATATGCTCAACCTCGGCGCAAACTACCGCGCAATACCGCGCTATGGCGAGTGG 900
 Qy 720 AsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPhe 739
 Db 901 AATTTGGTACGTCAAAGCGCAACACTGCTCAACCAATCCGTTTACGCCACAGAGCTTC 960
 Qy 740 LeuSerAspThrProGlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
 Db 961 CTCTCTGATACGCCGCAAAATGGCGCGAGCTTTACCGCGCGGTGACGTAAGTTT 1017

RESULT 15

XX Moraxella catarrhalis BASB107 DNA.
XX BASB107; infection; pneumonia; otitis media; sinusitis;
KW therapy; diagnosis; vaccine; genetic immunization; antibacterial;
KW antibiotic; receptor; ss.
XX
OS Moraxella catarrhalis.
XX
XX WO2000071724-A2.
XX
XX 30-NOV-2000.
XX
XX 18-MAY-2000; 2000WO-EP04618.
XX
XX 24-MAY-1999; 99GB-0012038.
XX 24-MAY-1999; 99GB-0012040.
XX 01-JUN-1999; 99GB-0012674.
XX 01-JUN-1999; 99GB-0012705.
XX 02-JUN-1999; 99GB-0012838.
XX 08-JUN-1999; 99GB-0013354.
XX
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Thonnard J;
XX
XX WPI; 2001-025166/03.
XX P-PSDB; AAB19945.
XX
XX New BASB103-108 polypeptides isolated from Moraxella catarrhalis
PT bacterium, useful for diagnosing and producing vaccines against
PT bacterial infections such as otitis media and pneumonia
XX
XX Claim 8; Page 73-74; 79pp; English.
XX

CC The present sequence is that of Moraxella catarrhalis ATCC 43617
CC DNA encoding BASB107 (see AAB19945), a protein with amino acid
CC sequence homology to Escherichia coli FhuE receptor precursor
CC outer membrane receptor for Fe(III)-coprogen,
CC Fe(III)-ferritoxamine-B and Fe(III)-rhodotorulic acid. The
CC invention provides M. catarrhalis BASB103-108 polypeptides (see
CC AAB19941-46) and polynucleotides (see AAA89212-17), and methods for
CC producing the polypeptides by recombinant methods. Claimed vaccine
CC compositions comprise a BASB103-108 polypeptide, or a BASB103-108
CC polynucleotide for genetic immunization. A claimed method for
CC diagnosing a M. catarrhalis infection involves identifying a
CC BASB103-108 polypeptide, or an antibody immunospecific to a
CC BASB103-108 polypeptide, in a biological sample. Compositions
CC comprising an immunologically effective amount of a BASB103-108
CC polypeptide, or a polynucleotide encoding it, are used to generate
CC an immune response in an animal. An antibody directed against a
CC BASB103-108 polypeptide can be used to treat humans with M.
CC catarrhalis disease. The polynucleotides may also be used to
CC generate probes and primers, as research reagents for the
CC discovery of therapeutics and diagnostics such as antibacterial
CC agents, and in the design of antisense sequences.
XX

SQ Sequence 2457 BP; 744 A; 560 C; 554 G; 599 T; 0 other;

Alignment Scores:
Pred. No.: 2,87e-132 Length: 2457
Score: 1585.00 Matches: 338
Percent Similarity: 58.92% Conservative: 141
Best Local Similarity: 41.57% Mismatches: 272
Query Match: 39.27% Indels: 63
DB: 22 Gaps: 13

US-09-936-377-2 (1-758) x AAA89216 (1-2457)

Qy 7 LysProIleValLeuSerIleIeu--LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db 22 AAACCGCTTGCTGGCCATATGGCAACTTTTCAATGTCATGCGAGCGGAAT 81

Qy 26 ---GluThrGluGlnSerValGlyLeuThrValThrValValGlyLysSerArgPro 44
Db 82 TTAAGGATAAGCAACCGTCATTTAGATGGCTTCGATCACCTCTTTAGTACCAA 141
Qy 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
Db 142 AATACAGATTGGGTTAATCATTTCAAAACAGTCAGTGCATCACAGTTTCAAAGAG 201
Qy 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
Db 202 CAACCTACAACCAACGACCACTAGCGATGCTTGGCAGGTGAGCTTGGCGTTTCA 261
Qy 85 AlaSerGlnTyrGlyGlyValAspAlaProValIleArgGlyGlnThrGlyArg 104
Db 262 TCTAACCATTTTGGGGCGGTGCTCAGCCCCCATCTTCTGGCGAGAGGTTAAACGC 321
Qy 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
Db 322 CTGAAATCTTACAAACGGTTTCAGAGGTTCTGGACATGTCTGGGTTCTGCCAGACCAT 381
Qy 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
Db 382 GCCATAGCGGTGGACACCACTGGCAAAACAGGTGGAGATTGTGCGAGGCTCTGGTGC 441
Qy 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
Db 442 TTGTGTACGCTCTGGCACTCAGCAGCGGTGGTCAATGCTGTGATGACAAAATACCC 501
Qy 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
Db 502 AGCAATTTGCCAGCAAA---TTACAGGTGATGTGACGGTGGCTTTAGCAGTCCCAAC 558
Qy 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
Db 559 CGTGAATAATTAATCAGCGCAGTGGCAGAGCCCACTGGGAGAGCATGTGCGAGTGGCT 618
Qy 205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
Db 619 GTTGCAGGCTGTCCAAACAGCAGCAGACTATAAAACGCCACGCTTTTGACGCGCATGTC 678
Qy 221 -----ArgAsnLeu 223
Db 679 TTTAACAAAAACATGAGATGATAACACTCAGCCAGCAATTCATCTATAAGACACCTTA 738
Qy 224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
Db 739 AAGCATCTGCCAGACAGCCATGCCAAATCAAACGCGAGCAAGCT-TGGCGTGTCTAGGT 797
Qy 244 TrpArgLysArg-PheTyrArgArgThrTyr-SerAspArgArgAspGlnTyrGlyLeuPr 263
Db 798 TGGCAATCAAGCTTTTGGGGGATCGTGAGCTTACGCCGAGCAATATGGCGCTGCC 857
Qy 263 oAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuI 283
Db 858 CAACCATTCATGATGAATGAAGATGTAGCGTGCATGGCAATTTCTAGTCCGCTTACA 917
Qy 283 eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAs 303
Db 918 ATACAAGCCATATTTGGCTTTGTATCTCTTTTATGGAATAATGATGATGATGATG 977
Qy 303 PAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAspAlaHis 318
Db 978 CAATGCGGGCTTGAATGCCATACATGATGACCAACGACCAACGACGACGACGACGAC 1037
Qy 318 sAlaHisAlaHis-----AsnGlyLysProTrpIleAspLeuArgAsnLys 333
Db 1038 TGACCACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1097
Qy 333 sArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgVa 353
Db 1098 GCGTTATGATGCAAGGCGCAATCAATCGCCGCTTTGCTGGCATTTGATAAATCCGAGC 1157
Qy 353 lHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPh 373

```

Db 1158 CAGCATGGGCAAGTGGATTATCATATGATGAGATAGATGGGGGTGAGAAGACCAAGCTT 1217
QY ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPhe 390
Db 1218 TTTTGATAATCAAGCAATGTGGCGTCTGGAAGCCTCACATACCCCATTCATAGCC 1277
QY 391 ----GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
Db 1278 GATGGGCAAGTTTACGGGGGTGTTTGGGTAGGTATCTCACCTCAAAAACACGCGACT 1337
QY 409 u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
Db 1338 TGTGCCACCTCGTTATGACATGCGCAATAAACAAGACACCCCAAAACATCTTGCACAATAA 1397
QY 424 nlysValGlnHisTyrSerPheGlyValGluGlnAlaAsn---TrpAspAsnPheTh 443
Db 1398 TAAACCAAAACAGGCGAGTGTGTTTGGTTTGAAGAATACAAAACCAACCAATGACAAGCTGAC 1457
QY 443 rLeuGluGlyGlyValArgValGluGlnLysAlaSerIleArgTyrAspLysAlaLe 463
Db 1458 CGTGTGACGGCGCGCTCGCATTGAGAAACAACCATCATCGATGATTATGATAAAGCGC 1517
QY 463 uille-----AspArgGluAsnTyrTyrLys 471
Db 1518 CATTTATCAGAGCTTAACTTAGCTTAGCAACGGCTCATGAACCAAGACATACGCTTTAA 1577
QY 471 sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
Db 1578 AGCATTTGCTGACAGCGGTACTTTAAACCCCAAAACCAACCGCACGCTCTTATGCTGT 1637
QY 489 uSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl 509
Db 1638 TGGGAGCATTTACAATTAACGCCCAACAATAAATATGCTGAATCTGTCGATCAAGA 1697
QY 509 uArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
Db 1698 AGCCTGCGCAATGCTCAGGAATGTATGCTACGCGCATGCCCTTGCCACCAACTCGTT 1757
QY 529 eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549
Db 1758 TGAATTTGGCAACCGCTTTTAAACAAAGAAATAATCAACATTCATTTGGGGCTGAC 1817
QY 549 yTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
Db 1818 ATTTCAAGGTGATAATGAGTATGCTCTGGGGCTATCATATGATTTTGATAACTA 1877
QY 569 rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerIleGluAspAspSe 588
Db 1878 TGTGTTTTTACAAACATTTGCGCAGTATAAGCAAGGTTTGGCGGCATGCGTCATGATAA 1937
QY 588 rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGl 608
Db 1938 AGACTTAAACACCGCAGCTATGAACAGCAGCGCGCAAAATTTATGGATTGATGATCA 1997
QY 608 uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGl 628
Db 1998 CATCGGTATCAGATTAATGATGATATATCATGCGCGTATTGCGGATTAATTCGTGG 2057
QY 628 yArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPh 648
Db 2058 CAAGCTCACCAATTTCCCTGACAAAAAGGCGAGAACCGATGCGTATGGCAACCGTCTCT 2117
QY 648 eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668
Db 2118 CATCAACAGCCAGCAGTCATACGCAAGCTGCGCAACCAACCAACCGCTTGGCATGAAT 2177
QY 668 uLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGl 688
Db 2178 AACCGCAATGTTAATGCAAAATGGTCAGGGTTTTTGGATATATCGCCATACCTTTAAACA 2237
QY 688 nAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAl 708

```

```

Db 2238 AGATAAATGGCGAATTTTGAACGCCCAACCCAGCTCATACTTGGTGAATTTGGGGCT 2297
QY 708 aAsnTyrArgArgAsnThrArgTyr-----GlyGluTyrAsnTyrTyrValLysAlaAs 726
Db 2298 TAATCATCAGCACAAAGCCAGCCATCAAGCAGGCTCGGTTCAAGTATTTTTTAATGCTAA 2357
QY 726 pAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMe 746
Db 2358 CAATCTATTAAACGATAAAGTCTTGTCTCATGAGACATTTTCCAGACATGCCACAAAT 2417
QY 746 tGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2418 GGGCGCAAACTTTATGCTCGGGGCAAAATTTCAAATTT 2454

RESULT 17
AAL46518
ID AAL46518 standard; DNA; 2646 BP.
XX
AC AAL46518;
XX
DT 19-JUL-2002 (first entry)
XX
DE M catarrhalis MCA102062 gene SEQ ID NO: 51.
XX
KW Moraxella; vaccine; respiratory tract infection; antiinflammatory;
KW auditory; antibacterial; otitis media; sinusitis; pneumonia; gene; ds.
XX
OS Moraxella catarrhalis.
PN WO200218595-A2.
XX
PD 07-MAR-2002.
XX
PF 28-AUG-2001; 2001WO-CA01221.
XX
PR 28-AUG-2000; 2000US-228294P.
PR 28-AUG-2000; 2000US-228295P.
PR 28-AUG-2000; 2000US-228296P.
PR 28-AUG-2000; 2000US-22838P.
PR 29-AUG-2000; 2000US-228439P.
PR 29-AUG-2000; 2000US-228441P.
PR 29-AUG-2000; 2000US-228442P.
PR 29-AUG-2000; 2000US-228443P.
PR 29-AUG-2000; 2000US-228511P.
PR 29-AUG-2000; 2000US-228512P.
PR 29-AUG-2000; 2000US-228742P.
PR 29-AUG-2000; 2000US-228773P.
PR 01-SEP-2000; 2000US-229465P.
PR 01-SEP-2000; 2000US-229474P.
PR 01-SEP-2000; 2000US-229475P.
PR 01-SEP-2000; 2000US-229478P.
PR 05-SEP-2000; 2000US-229740P.
PR 05-SEP-2000; 2000US-229803P.
PR 05-SEP-2000; 2000US-229804P.
PR 05-SEP-2000; 2000US-229805P.
PR 05-SEP-2000; 2000US-229806P.
PR 05-SEP-2000; 2000US-229809P.
PR 06-SEP-2000; 2000US-229811P.
PR 06-SEP-2000; 2000US-230214P.
PR 06-SEP-2000; 2000US-230250P.
PR 06-SEP-2000; 2000US-230252P.
XX
PA (AVET ) AVENTIS PASTEUR LTD.
XX
PI Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;
XX
DR WPI; 2002-401721/43.
DR P-PSDB; AAO17586.
XX
PT Moraxella polypeptide and polynucleotides useful as vaccine for
PT immunizing a host e.g. humans against disease e.g. otitis media,
PT pneumonia, caused by infection of the bacteria -

```

XX Claim 2; Fig 51; 277pp; English.
 XX The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarrhalis. These can be used to produce
 CC vaccines which protect against M. catarrhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a coding sequence of the invention.
 XX
 SQ Sequence 2646 BP; 817 A; 583 C; 571 G; 675 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,18e-132 Length: 2646
 Score: 1585.00 Matches: 338
 Percent Similarity: 58.92% Conservative: 141
 Best Local Similarity: 41.57% Mismatches: 63
 Query Match: 39.27% Indels: 63
 DB: 24 Gaps: 13
 US-09-936-377-2 (1-758) x AAL46518 (1-2646)
 QY 7 LysProIleValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
 DB 110 AAACCGCTTGGCGTGGCCATATGGCAACTTTTCAATGCCAATGCTGGCAGAGCGAAT 169
 QY 26 ---GluThrGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
 DB 170 TTAAGGATAACCAACCGCTCATTTAGATGGCGTTTCGATCACTCTTTAGCTGACCAA 229
 QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
 DB 230 AATACAGAGTTGGCGTTAATCATTCATAAAACAGTCAGTGGCATCACAGTTTCAAAGAG 289
 QY 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
 DB 290 CAACATACACACAGACCAACCCCTAGGCGATGCCCTTGGCAGAGCTTGGCGTTTCAT 349
 QY 85 AlaSerGlnThrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
 DB 350 TCTAACCATTTGGGGCGGTGCTTCAGCCGCCATCATTCCTGGGGCAGAGGTTAAACGC 409
 QY 105 IleLysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 DB 410 CTGAAATCTCAAAACCGTTGAGAGGTGGGACATGCTCGGTGGTGGCCAGACCAT 469
 QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
 DB 470 GCCATAGCGGTGGACACCACTGGCAAAACAGGTGGAGATTGTGCGAGGCTCTGCTGCC 529
 QY 145 LeuLeuThrSerSerGlyValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
 DB 530 TTGTTGTAGCCTCTGGCAACTCAGACGGCGTGGTCAATGCTGTTGATGACAAAATACC 589
 QY 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
 DB 590 AGCAAAATGGCCAGCAAA---TTACAAGGTGATGTGACGGTGGCTTTAGCAGTCCCAAC 646
 QY 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
 DB 647 CGTGAATAATTAATCCCGCAGTCCGAGCCGCCACTGGGAGAGCATGTGCGAGTGGGT 706
 QY 205 ThrGluGlyLeuThrArgLysSerGlyAspThrAlaValProArgTyr----- 220
 DB 707 GTTGACGGGCTGTCCAACAGCAGCAGCACTATAAAACGCCACGCTTTGACCGCCATGTC 766
 QY 221 -----ArgAsnLeu 223
 DB 767 TTTAAACAAAAACATGAAGATGATAAACAACAGCAAGATTCATCTATAAAGACACCTTA 826
 QY 224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
 DB 827 AAGCATCTGCCAGACAGCATGTCCTCAATCAAAACCGAGGAACGCT-TGGCGTGTCTATGGT 885

QY 244 TrpArgLysArg-PheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuPr 263
 DB 886 TGCATCATCAAGGCTTTTGGGGCATCGGTAGCTTACGCCGAGACAAATATGCGCTGCC 945
 QY 263 oAlaHisSerHisGluTyrAspCysHisAlaAspIleIleTrpGlnLysSerLeuI 283
 DB 946 CAACCATTCACATGAATATGAAGATGTAGCGTGCATGTGGCAATTTCTCAGTCCGCCCTTACA 1005
 QY 283 eAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuGluGluAspValAspTyrAs 303
 DB 1006 ATACAGCCATATTTGCGTTTGTATCTTTTGTATGGAAATATGATGACTTAGAGTTGA 1065
 QY 303 pAsnProGlyLeuSerCysGlyPheHisAspAsp-----AspAlaHis 318
 DB 1066 CAATGCGGGCTTGAATGCCATACATGATGACACAGCAGCAGCAGCAGCAGCATGTCTCA 1125
 QY 318 sAlaHisAlaHis-----AsnGlyLysProTrpIleAspLeuArgAsnLys 333
 DB 1126 TGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1185
 QY 333 sArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgVa 353
 DB 1186 CGCTTATGATGTCAAGGCAAAATCAATCGCGCTTTGCTGGCATGTATAAAATCGCAGC 1245
 QY 353 lHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAlaValGluAsnPh 373
 DB 1246 CAGCATGGCAAAATGGATTATCATCATGATGATAGTAGTGGGGTGAGAGACCCAGCTT 1305
 QY 373 ePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle----- 390
 DB 1306 TTTTGAATATCAAGCAATGTGTGGCGTCTGGAAGCCTCACATACCCCATTCATACGCC 1365
 QY 391 ----GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLe 409
 DB 1366 GATGGCGAAGTTTAGCGGGTGTGGGTAGGTATCTCACCCTCAAAAACAGCGGACT 1425
 QY 409 u-----SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAs 424
 DB 1426 TGTGCCACCTCGTTATAGATGGCAATAAACAAGACACCCCAAAACATCTTGCAATAA 1485
 QY 424 nLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn---TrpAspAsnPheTh 443
 DB 1486 TAAACCAAAACAGCGAGTGTGTTTGGTTTGAAGAATAACAACCAACCAATGCAAGCTGAC 1545
 QY 443 rLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLe 463
 DB 1546 CGTTGACGCGCGCTCGCATTTGAGAAACAACCATCACCATGGATTATGATTAAGACGC 1605
 QY 463 uIle-----AspArgGluAsnTyrTyrLys 471
 DB 1606 CATTTATCAGAGCTTAAACTTAGCGCTTAGCAACCGCTCATGAACCCAGACATACGCTTAA 1665
 QY 471 sGlnProLeuProAsp-----LeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLe 489
 DB 1666 AGATTGTGACACCGGTACTTTAAACCCCAAAACAACCAACCGCGCTCTTATGCTGT 1725
 QY 489 uSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGl 509
 DB 1726 TGGGACGCAATTACAAATAACGCCCAACATAAATATCGCTGAATCTGTCGCATCAAGA 1785
 QY 509 uArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPh 529
 DB 1786 ACGCTGTGCAAAATGTCTAGGAATTTGTATGCTCAGCGCATGCACCTTGCACCACTCGTT 1845
 QY 529 eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549
 DB 1846 TGAATTTGGCAACCGCTTTTAAACAAAGAAATAATCCAACAACATTTGATTTGGGCTGAC 1905
 QY 549 yTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
 DB 1906 ATTTCAAGTGTATAAATGGGATTATCGTCTTGGGGGCTATCATTTATGATTTTGTAACTA 1965
 QY 569 rIleTyrAlaGlnThrLeuAsnAsp---GlyArgGlyProLysSerIleGluAspAspSe 588

```

Db 1966 TGGTGTTCACAACTGCGCAGTATAGACAGGTTTCGGTGGCATCGTCAATATAA 2025
Qy 588 rGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGI 608
Db 2026 AGACTTAAAAACCGCAGCTATGAACAGCAGCGGCGGAAATTTTATGGATTTGATGCA 2085
Qy 608 uileTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGI 628
Db 2086 CATCGGTTATCAGATTAAATGATGATATCATGTGGCGTTATTTGGTGATTAATTCGTGG 2145
Qy 628 YArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProph 648
Db 2146 CAAGCTCACCAATTTGCTGACAAAAAGGCGAGAACCGATGCGTATGGCAACCGTCTCT 2205
Qy 648 eileAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668
Db 2206 CATCAAAACAGCAGCAGTATAGCCCAAGACTGCCAACCAACCGCTTGGCATGAAAT 2265
Qy 668 uLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGI 688
Db 2266 AACCGCCAATGTTAATGCAAAATGCTCAGGCTTTTGGATATCGCCATACCTTTTAAACA 2325
Qy 688 nAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAl 708
Db 2326 AGATAAATTCGCGAAATTTGAACGCCCAACCCAGCTCATAACTTGGTGAATTTGGGGCT 2385
Qy 708 aAsnTyrArgArgAsnThrArgTyr-----GlyGluTyrAsnTyrValLysAlaAs 726
Db 2386 TAACATCAGCACCAAGCCAGCCATCAAGCAGGCTCGTTTCAGGTATTTTAAATGCTAA 2445
Qy 726 pAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMe 746
Db 2446 CAATCTATTAACCATAAAGTCTTTGCTCATGACACATTTTCCAGACATGCCACAAAT 2505
Qy 746 tGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2506 GGGCGGAAACTTTATGCTCGGGCAAAATTTCAAATTT 2542

RESULT 18
AAF28547
ID AAF28547 standard; DNA; 89047 BP.
XX
XX
AC AAF28547;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #34.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
FN WO200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
XX
XX WPI; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX

```

```

PS
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localized infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 89047 BP; 26501 A; 17338 C; 19060 G; 26147 T; 1 other;

```

```

Alignment Scores:
Pred. No.: 4,06e-130 Length: 89047
Score: 1585.00 Matches: 338
Percent Similarity: 59.92% Conservative: 141
Best Local Similarity: 41.57% Mismatches: 272
Query Match: 39.27% Indels: 63
DB: 22 Gaps: 13

US-09-936-377-2 (1-758) x AAF28547 (1-89047)
Qy 7 LysProIleValLeuSerIleLeu---LeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db 424 AAACCGCTTGGCTGTGCCATATTGGCAACTTTTTCATGCAATGCTGCGAGAGCGAAT 483
Qy 26 ---GluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
Db 484 TTAAGAGTAAACCCACCGTCATTTTAGATGCGGTTTCGATCACCCTTTTAGTGACCA 543
Qy 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
Db 544 AATACAGAGTTTGGCGTTAATCAATCAAAAACAGTCAGTGGCATCACAGTTTCAAAAGAG 603
Qy 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
Db 604 CAACTACACAAACAGCAGCAACCCCTAGCGCATGCTTGGCAGGTGAGCTTGGCGCTTCA 663
Qy 85 AlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
Db 664 TCTAACCATTTTGGGGGCGGTGCTTCAGCCCCCATCATCTTCGTGGCAGAGGTAAACGC 723
Qy 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
Db 724 CTGAAATCTACAAAACCGTTTCAGAGGTTTGGACATGCTCGGTTGTGCGCAGACCAT 783
Qy 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
Db 784 GCGATAGCGTGGACACCACTGGCAAAACAGGTGGAGATTGTGCGAGGCTCTCGTGCC 843
Qy 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
Db 844 TTGTTGTACGCTCTGGCAACTCAGCAGCGGTGTCATGTCTGTTGATGACAAATATACC 903
Qy 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
Db 904 AGCAATATGCCAGCAAAA---TTACAAGGTGATGTAGCGGTCTTTAGCAGTGGCAAC 960
Qy 185 LeuGluLysLeuThrSerGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
Db 961 CGTGAAAATTAATCACCGCCAGTCCGAGCCCACTGGGAGAGATGTGGCAGTGGCT 1020
Qy 205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
Db 1021 GTTGAGGCGCTGTCCAAACAGCAGCAGACTATAAAACGCCACGCTTTGACCGCCATGTC 1080
Qy 221 -----ArgAsnLeu 223
Db 1081 TTTAACAAAAAATGAAGATGATAACACTAGCCAGCAATTCATCTATATAAGACACCTTA 1140

```

Claim 1; Page 324-345; 545bp; English.

PR 29-AUG-2000; 2000US-228511P.
 PR 29-AUG-2000; 2000US-228512P.
 PR 29-AUG-2000; 2000US-228742P.
 PR 29-AUG-2000; 2000US-228773P.
 PR 01-SEP-2000; 2000US-229465P.
 PR 01-SEP-2000; 2000US-229474P.
 PR 01-SEP-2000; 2000US-229475P.
 PR 01-SEP-2000; 2000US-229478P.
 PR 05-SEP-2000; 2000US-229740P.
 PR 05-SEP-2000; 2000US-229803P.
 PR 05-SEP-2000; 2000US-229804P.
 PR 05-SEP-2000; 2000US-229805P.
 PR 05-SEP-2000; 2000US-229806P.
 PR 05-SEP-2000; 2000US-229809P.
 PR 05-SEP-2000; 2000US-229811P.
 PR 06-SEP-2000; 2000US-230214P.
 PR 06-SEP-2000; 2000US-230250P.
 PR 06-SEP-2000; 2000US-230252P.
 XX
 PA (AVET) AVENTIS PASTEUR LTD.
 XX
 PI Loomsore S, Wang J, Bradley B, Ochs M, Yang Y;
 XX
 DR WPI: 2002-401721/43.
 DR P-PSDB; AAO17585.
 XX
 PT Moraxella polypeptide and polynucleotides useful as vaccine for
 PT immunizing a host e.g. humans against disease e.g. otitis media,
 PT pneumonia, caused by infection of the bacteria
 XX
 PS Claim 2; Fig 49; 277pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarrhalis. These can be used to produce
 CC vaccines which protect against M. catarrhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a coding sequence of the invention.
 XX
 SQ Sequence 2942 BP; 911 A; 562 C; 632 G; 837 T; 0 other;
 Alignment Scores:
 Pred. No.: 4-53e-98 Length: 2942
 Score: 1205.50 Matches: 290
 Percent Similarity: 46.25% Conservative: 142
 Best Local Similarity: 31.05% Mismatches: 303
 Query Match: 29.87% Indels: 199
 DB: 24 Gaps: 18
 US-09-936-377-2 (1-758) x AAL46517 (1-2942)
 QY 9 IleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGlu 28
 DB 83 ATCATAACCGAGGTATCATGATAAAACCACTTGTGTGCGATATCGGCCACTTT 142
 QY 29 GlnSerValcylLeuGluThrValThrValValcylLeuSerArgProArgAlaThrSer 48
 DB 143 GCGATCCACCGTAGCAGATATACCAAGCTGGTGAGAGCAACCAACCCTTAAG 202
 QY 49 GlyLeuLeuHisThrSer-----
 DB 203 GGTGTATTGTAGCTCGCAACGAACCAAAATACAGGTTTGTATCTAATGATTCAAA 262
 QY 55 ThrAlaSerAspIstlelleSerGlyAspThrLeuArgGlnIstleValAlaSerLeuGly 74
 DB 263 CAATCAGTATCTTACGCTTTTCAAAAGATATAAATAATATATCGTTCCGCAACCTTGGGC 322
 QY 75 AspAlaLeuAspGlyValProGlyIleHisAlaSerGlnThrGlyGlyAlaSerAla 94
 DB 323 AATGCGTTAAGTGGTGGCTTGGTATTATAGTAACCTTTTGGTGGCGTTTATCTGCA 382
 QY 95 ProValIleArgGlyGlnThrGlyArgGlyIleValLeuAsnHisGlyGluThr 114
 DB 383 CCTGTTGCGAGGGAAGAGGGTGTGCGTCTTAAGATTTCACAAATGCAACTGATG 442

QY 115 GlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGln 134
 DB 443 ATTGATGTGTCATCAATATCGCTGATCATGTTGTGGCGACCATACACTTTAGCTCT 502
 QY 135 GlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGly 154
 DB 503 AAAAGTTGAGCTTGTGCTGCTGATACGCTGTTATATATGCTGGCTGGCATCGCCAGCTGGT 562
 QY 155 LeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly----- 171
 DB 563 GTGATTAATGTTGTTGATGACCTATCCGAATCGTATGCTAGTGGTGTATCCATGAC 622
 QY 172 ---ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSer 190
 DB 623 AAAATCGAAGCGAGAGCATGCTTCGATATAACACAAACACCATGAAAAGCTTGCAACT 682
 QY 191 GlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArg 210
 DB 683 GCAGGGGTGAGCTTTGGGGTAGGATCGCATTCGCTTCGGGTGGAGGGCTTAAAGCGA 742
 QY 211 LysSerGlyAspTyrAlaValProArgTyr-----ArgAsnLeuLysArgLeuPro 227
 DB 743 GAGGTGATGACTATCAAGTTCCTCCCATTTTCAGGCGAGATCGCATGTAGATTATGTCCA 802
 QY 228 AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly----- 243
 DB 803 GGTAGT-----GCAATAACTCTACCGTTGGCATGATTGGCGTCTTATATT 850
 QY 244 TrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPro 263
 DB 851 CATGATAATGGGCATATCGGTGCTTCTTATAGCCACCGTAAAGATCGTTATGTTATCCCA 910
 QY 264 AlaHisSerHisGlnTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIle 283
 DB 911 GGCATATCCACTGCGACGCCAACGAGAGCATTTTATCAATGGCATATATACACAAA 970
 QY 284 AsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLeuAspValAspTyrAsp 303
 DB 971 TCCCAATTATTTTACCCTTATCTCTTATGATGGAGGATTCAGATTATGAT---GAC 1027
 QY 304 AsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsn 323
 DB 1028 AATCTCTCATACGATTCGCCCAACACCAAGAGCATATCGGTGAGCATATAATCCCAAG 1087
 QY 324 Gly-----LysProTyrIleAspLeuArgAsnLysArgTyr 335
 DB 1088 GCGTGCCTCATCATCATGACATCATTCGCTTGGATTGATGATAAAACCAATTCGCTAC 1147
 QY 336 GluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu 355
 DB 1148 GACATTCGTGGCGAGGTATATCGGCTTATTCAGGTTTGGATAAAATTAAGCTTAAGCTTA 1207
 QY 356 AsnArgAspAspTyrHisHisAspGluLys---AlaGlyAspAla----- 369
 DB 1208 ACTATGCGAGATTATTTATCATGATGAAAAGATGCTGGCAATGAGCAAGCAACCAAT 1267
 QY 370 -----ValGluAsnPhePheAsn 375
 DB 1268 CACAAACCTTCTGAGCGTGATACACGGTGGATAAGGGTCATGCCAGCTCTATTATTTACA 1327
 QY 376 AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly 395
 DB 1328 AAAAAAGGCGTTAATGCTGCTTGGAGTTATATCATACACCGACCAACCGTTATCTGGG 1387
 QY 396 SerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAla----- 411
 DB 1388 GTATTGGGTATTGGGTATCAAAACCAAAATCTGCAGCAGGAGCGGTATTGTCACAGC 1447
 QY 412 -----ThrSerGluAlaValLysGlnPro----- 419
 DB 1448 TATTTTCAATCAGAGCAGATGGCAAAAGCCCAAGTCAAAACATTAAACCAATATCGT 1507

QY 420 -----MetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlnAla 437
 Db 1508 CTTTACTTATTAGTTCCTTCAATAAATAAAGCCCTTGGTATTTTGGACATTGACAACTA 1567
 QY 438 AsnTrpAspAsnThrLeuGluGlyValArgValGluLysGlnLysAlaSerIle 457
 Db 1568 AAGTAATCAATGACTTTTAAAGTGGCGATGGTCTATGAACACAAACACCAAT 1627
 QY 458 ArgTyrAspLysAlaLeuLeuAspArgGluAsnTyrTyr----- 470
 Db 1628 GAATATGACGACGACTTACTTGACCATGCTTTGACGATTTTAAAGTAAAGCAGCTA 1687
 QY 471 LysGlnPro---LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
 Db 1688 AAAGCCTGATCATCTGATTTGACGACATATAAACAACATGCCCTCTTTATGCTGGT 1747
 QY 490 SerGlyAsnTrpThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu 509
 Db 1748 AGTCCTTATGGATATTACGCCAATCATCGATTGTCATTGACTCTACTCATACGAA 1807
 QY 510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPhe 529
 Db 1808 CGCATTCATCGCGATGGAGCTGTATTATCAAGCGGACATTTGGCGACGCTCTTT 1867
 QY 530 GluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly 549
 Db 1868 GAGATGCAATAAACTTGGTCAAGAAATCGGATAATTATGACTGGGTTTTATG 1927
 QY 550 TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
 Db 1928 CATACAGCAGATAAAGTCAGCTATAAAGCAGCACTTACTATAGCAATTTTGATACTAT 1987
 QY 570 IleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGlu 589
 Db 1988 ATTTTAAATGAGACC-----ATTGCCAAAGAGGCAAT 2020
 QY 590 MetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIle 609
 Db 2021 TTATACATCAGACGCTAATAACACAGCGGTAAAGTTTATGGTGGAGGGTTCAATTA 2080
 QY 610 TyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArg 629
 Db 2081 ACTTACCAGCCAAATGCCAATCACAGTGTGTGTTTTTTGGTGATATGTCAGAGTAAA 2140
 QY 630 LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
 Db 2141 ATTGGTGCAATTATCTGATATTAAAGGTAAGCTT---GTATATGCTGGCAGAAAGTGGTT 2197
 QY 650 AlaGlnAspAspGln----- 654
 Db 2198 TATTTGATGATATATCAAGGATATGACCGTTGATGATATGTTGATATGATGCCGAT 2257
 QY 654 ----- 654
 Db 2258 GGTGGTTTGACTGTGCTTAAACAGCGCTGAACAATGGGACAAATTAACGATAATAAT 2317
 QY 654 ----- 654
 Db 2318 GATTGTAGTACACCAATTAATGTCTATAAAACGGTACAAACCTCAGCGAGGAGGAT 2377
 QY 655 -----AsnAlaProArgValProAlaAlaArgLeuGly 665
 Db 2378 TACGACCGTTTGGCAGCAATCCAACTTACGACCCAGAGTACCGCCCGCTTGGGC 2437
 QY 666 PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
 Db 2438 ATTCGTTGGCAAGGGTATTTTGGTGTGCTGCAATGCAAGATTTTAACCATGTG 2497
 QY 686 PheAlaGlnAsnLysLeuAla----- 692
 Db 2498 TTTGCACAAAATAAGTTGGCCACCTCAACGGTTGCCATTAAACCTCAATTCAAGCAGCCA 2557
 QY 692 ----- 692

Db 2558 GAAGTTGCCAACCCCATGAGAGTCATTGCCGAATCAGCGACTATGGCAGTGATAACAAC 2617
 QY 693 -----ArgTyr-----GluThrArgThrProGlyHisHisMetLeu 704
 Db 2618 CCTTTGATGATGACCCCAAGATATATACAGAAAACAAACGGCAGGATATAATTGCTC 2677
 QY 705 AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLys 724
 Db 2678 AATGTTGGCTTAGATATTAAACAATGCATATCGTAATGTTGATTATACGCTGTCAATCGT 2737
 QY 725 AlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrPro 744
 Db 2738 GCGAATAATTTACTTAATGAACAATCTACATTCAACAATCATTTTTCGCTTGTACCG 2797
 QY 745 GlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2798 CAGATGGCGCTAATCTGACITTAGCTTTGACGGCTAAATTT 2839
 RESULT 20
 AAF28554
 ID AAF28554 standard; DNA; 269223 BP.
 AC AAF28554;
 XX 04-APR-2001 (first entry)
 DE Genomic fragment #41.
 KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.
 OS Moraxella catarrhalis.
 XX WO200078968-A2.
 XX 28-DEC-2000.
 XX 16-JUN-2000; 2000WO-US16649.
 XX 18-JUN-1999; 99US-0140121.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Lagace RE, Patterson C, Berg KL;
 WPI; 2001-041427/05.
 Genomic library for identifying diagnostic and therapeutic
 compositions, and for identifying virulence factors, regulatory
 elements and drug targets, comprises Moraxella catarrhalis nucleic
 acids -
 Claim 1; Page 486-545; 545pp; English.
 The present invention relates to a Moraxella catarrhalis genomic library
 comprising of a combination of 41 nucleic acid molecules (see
 AAF28514-AAF28554). The library has a number of uses described in the
 specification e.g. is useful for identifying diagnostic and therapeutic
 compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 aerobic, gram-negative diplococcus, normally found among the bacterial
 flora of human upper airways. M. catarrhalis is known to cause acute,
 localised infections such as otitis media, sinusitis and bronchopulmonary
 infection and life-threatening, systemic diseases including endocarditis
 and meningitis.
 SQ Sequence 269223 BP; 77067 A; 56596 C; 57380 G; 78180 T; 0 other;
 Alignment Scores:
 Pred. NO.: 2.3e-95 Length: 269223
 Score: 1205.50 Matches: 290
 Percent Similarity: 46.25% Conservative: 142
 Best Local Similarity: 31.05% Mismatches: 303

[illegible]

1426 AATACCAATAAAAGCCTTGGTATTATTTTGGACTTGGACAACTAAAGCTAAATCAAAATGACT 1488

444 LeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeu 463

1486 TTTAAGTGGCGATGCGTCATGAAGACAAACCAACCAATTGAATATGACCAGCATTTA 1545

464 IleAspArgGluAsnTyrTyr-----LysGlnPro---LeuPro 475

1546 CTTGACCATGCTTGGCAGTATTTTTTAAGTAAAGCAGCAGCTAAAGCAGCTGATCATCCT 1605

476 AspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPhe 495

1606 GATTTCAGCAGCATATAAACACATGCCACTCTTATGCTGGTAGTGCCTTATGGGATATT 1665

496 ThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGln 515

1666 AGCCAAATCATGATGTGTCATTGACCTACTACATAAGCAACGATTCATTCGCGCGATG 1725

516 GluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHis 535

1726 GAGCTGTATTATCAAGCGCGACATTTGGCGCACCAGCTCTTTTGAGCATGGCAATAAAAC 1785

536 LeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrp 555

1786 TTGGTCAAGAAAATCGGATAATTATGAGTGGTTTTATGCATACAGCAGATAAAGTC 1845

556 GlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeu 575

1846 AGCTATAAGCAAGCACTTACTATAGCNAATTTTGATTAATATATTTTAAAGAGACC--- 1902

576 AsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr 595

1903 -----ATTGCCAAAGAGCAATTTTATCATCAGACGCTAT 1938

596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615

1939 AATCAGACGCGCTTAAGTTTTATGGTGGAGGGTTCAATTAACCTACACGCCAAATGCC 1998

616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635

1999 ATCAGCTGTGATGTTTTTGGTGATATGGTGCAGGTAAATTTGGTGCAATTATCTGAT 2058

636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGln--- 654

2059 ATTAAAGGTAAGCTT--GTATATGCTGGCAGAAAGTGGGTTATTTTGTATGATGATC 2115

654 ----- 654

2116 AAGGATATCAGCTGTGATGATAATGGTGATATTATGATGCCGATGGTGGTTGACTTGTGCC 2175

654 ----- 654

2176 TTAAAAAGCCTGAACAAATGGGACAAATTAACGATAATATGATTGTAGTACAAACCAT 2235

654 ----- 654

2236 AATGCTATAAAAAGGTACAAACCTCAGCGAGGAGGATTACGACCGTTTGGCAGCA 2295

655 -----AsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSer 671

2296 AATCAACTTACGACCCAGATACCGCCGCGCTTGGGCAATTCGTTGGCAAGGGTAT 2355

672 LeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeu 691

2356 TTTGGTGATCATTTGGTCTGCCAATGCGAAATTAACCATGTGTTTGCAAAATAAAGTT 2415

692 Ala----- 692

2416 GCCACTCAACGGTTGCCATTAAACCTCAATTCAACGACCGCAGAGGTTGCCACGCCAT 2475

692 ----- 692

2476 GAGAGCTATTGCCGAATCAGCGACTATGGCAGTGATAACAACCCCTTTGATGATGACGCA 2535

Qy 693 ArgTyr-----GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyr 710
 Db 2536 AGATATATACAGAAAAACAAACGCGAGATATAATTGCTCAATGTTGGCTTAGATTAT 2595
 Qy 711 ArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsn 730
 Db 2596 AACAAATGATATCGTAAATGTTGATTATACGCTGTCAATTCGTGCAATTAATTACTTAAT 2655
 Qy 731 GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPhe 750
 Db 2656 GAAACAATCTACATTCACAACACTCAATTTTGGCGTTGTACCGCAGATGGGGCGTAATCTG 2715
 Qy 751 ThrGlyGlyValAsnValLysPhe 758
 Db 2716 ACTTTAGGTTTGACGGCTAAATTT 2739

RESULT 22
 AAZ53317
 ID AAZ53317 standard; DNA; 522 BP.
 XX AC AAZ53317;
 XX DT 21-MAR-2000 (first entry)
 XX DE Neisseria gonorrhoeae ORF 147 partial DNA sequence SEQ ID NO:583.
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 XX KW antibacterial; gene therapy; ds.
 XX OS Neisseria gonorrhoeae.
 XX FN WO9957280-A2.
 XX PD 11-NOV-1999.
 XX PF 30-APR-1999; 99WO-US09346.
 XX PR 01-MAY-1998; 98US-0083758.
 XX PR 31-JUL-1998; 98US-0094869.
 XX PR 02-SEP-1998; 98US-0098994.
 XX PR 02-SEP-1998; 98US-0099062.
 XX PR 09-OCT-1998; 98US-0103749.
 XX PR 09-OCT-1998; 98US-0103794.
 XX PR 09-OCT-1998; 98US-0103796.
 XX PR 25-FEB-1999; 99US-0121528.
 XX PA (CHIR) CHIRON CORP.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 XX PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 XX PI Tettelin H, Venter JC;
 XX WPI: 2000-062150/05.
 XX P-FSDB; AAY74555.

Novel Neisserial polypeptides predicted to be useful antigens for
 vaccines and diagnostics -
 Claim 7; Page 412; 1453pp; English.
 AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
 PCR primers used in the exemplification of the present invention. The
 polypeptides, the polynucleotides, antibodies and compositions of
 the invention can be used as vaccines, as diagnostic reagents, and as
 immunogenic compositions. The polypeptides can be used in the
 manufacture of medicaments for treating or preventing infection due to
 Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX Sequence 522 BP; 126 A; 156 C; 135 G; 105 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 5.97e-63 Length: 522
 Score: 803.00 Matches: 163
 Percent Similarity: 98.81% Conservative: 3
 Best Local Similarity: 97.02% Mismatches: 1
 Query Match: 19.90% Indels: 1
 DB: 21 Gaps: 0

US-09-936-377-2 (1-758) x AAZ53317 (1-522)
 Qy 1 MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
 Db 19 ATGGCACAATCACAACCTCAAAACCCATTGTTTATCAATCTTTTAAATCAACACCCCTC 78
 Qy 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
 Db 79 CTGCCCCAAGCGCATGAAACTGAGCAATCGGTGGGCTTGGAAACGCTCAGCGCTCGCGC 138
 Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 Db 139 AAAAGCCGTCGCGCGCGACTTCGGGGCTGTGTCACACTTCGACCGCTTCGACAAAATC 198
 Qy 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyValAspAlaLeuAspGlyVal 80
 Db 199 ATCTCGGCGATACTTTGGGCCMAAAAGCCGTCACCTTGGGCGACGCTTTGGACGGCGTA 258
 Qy 81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGln 100
 Db 259 CCGGCGATCCAGCTTCGCAATACGCGCGCGCGCATCCGCTCCGCTTATTCCGCGTCAA 318
 Qy 101 ThrGlyArgArgIleLysValLeuAsnHisGlyGlyGluThrGlyAspMetAlaAspPhe 120
 Db 319 ACGGCGACAGCGATTAAAGTATTGAACCATCAGCGCAAAACGGCGATATGCGGACTTT 378
 Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
 Db 379 TCTCCGATCAGCGCATTTATGGTAGATACCGCTTGTGCAACACAGGTTGAATCTCGCG 438
 Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaIleGlyLeuValAspValAlaAs 160
 Db 439 GGGCGGTTAGCTCTTGTACAGCTCGGCAATGTGGCGGGCTGGTCAATGTTGCCGA 498
 Qy 160 pGlyLysIleProGluLysMet 167
 Db 499 TGGAAAAATCCCCCAAAAAATG 520

RESULT 23
 ABQ90153
 ID ABQ90153 standard; DNA; 2127 BP.
 XX AC ABQ90153;
 XX DT 01-OCT-2002 (first entry)
 XX DE M. capsulatus gene #138 for DNA array.
 XX KW Micro array; gene; ds; differential expression; gene expression.
 XX OS Methylococcus capsulatus.
 XX PN WO200255655-A2.
 XX PD 18-JUL-2002.
 XX PF 14-JAN-2002; 2002WO-NO00019.
 XX PR 12-JAN-2001; 2001NO-0000235.

QY 522 shisValalaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
 Db 1440 ACACGACCCACCCACAGCTACGAGCTGGCAATCCCAATCTCACCGAGGAACCTCTTA 1499
 QY 542 nAsnileGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTy 562
 Db 1500 CAACCTTCAGCTGGGCTACCGCTACAAGGGCGAGAGTCACGGCGAGCTCAACGTGT 1559
 QY 562 rArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProly 582
 Db 1560 CCAGAATTTCGTGAACGACTACATCTACAGCAGATCACGGATGAG----- 1605
 QY 582 sSerileGluAspAspSerGlu-----Me 590
 Db 1606 -GTGTTCAACGAAGATACCGAAGTTCGAGAAAGTGTGCACAATCCGGGGCGCTGTT 1664
 QY 590 tLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTy 610
 Db 1665 CCCGTGTGTGCAGACCACTACGCGCGGCGGTATTCTCTGGCTACGAAGGCAACGTGAG 1724
 QY 610 rPheLysProThrProArgTyrArgIleGlyVal-----SerGlyAspTy 625
 Db 1725 ATTT---CCGTGTGTGAAGAACCAATTACGCGCTGATCGACCTCGACCTGTTTCAGCGACT 1781
 QY 625 rValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAs 645
 Db 1782 CACCGCGCGCGAGTTCGTCAAC----- 1803
 QY 645 nArgProPheIleAlaGlnAspGlnAsnAlaProArgValProAlaAlaArgLeuG 665
 Db 1804 -----GGCGGCAACGTGCGCGCATGCGCGCTCGGCTACGG 1841
 QY 665 yPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVa 685
 Db 1842 CTTCCAGCTGACTATGCTGCGAGCAGCATGAGCGGCAATCTCGGCTACTCGCG 1901
 QY 685 lPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAs 705
 Db 1902 CGAGAGCAGATCTATGCGGCGCAGACGACGCCATGACCAACGGCTAGCTGCTCTCAA 1961
 QY 705 nLeuGlyAlaAsnTyrArg---ArgAsnThrArgTyrGlyGluTrpAsnTrpTyrVally 724
 Db 1962 CTTGGTGTGCAATACGAGTCAAGCATTCAGAGCGCGCAAGTCTGCTCTGCTCA 2021
 QY 724 sAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSer---AspTh 743
 Db 2022 AGGCAACAACCTGCTCAACAGACACCTCCGAAACTCCACTCTATCTCGCTATTTCGC 2081
 QY 743 rProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2082 GCCGAGGCGGCGCGGGGTGAGTTGGGCTACGATCACTAT 2127
 RESULT 24
 ID AAF26431
 XX AAF26431 standard; DNA; 1173 BP.
 AC AAF26431;
 XX AAF26431;
 DT 02-MAY-2001 (first entry)
 DE Pseudomonas sp heavy metal transporter encoding DNA ORF08372.
 KW Heavy metal transporter; iron transporter; transgenic plant;
 KW homeostasis regulator; heavy metal ion; trace element; soil pollution;
 KW plant growth promoter; plant development; ds.
 OS Pseudomonas sp.
 PN DE19934720-A1.
 XX 25-JAN-2001.
 PD 99DE-1034720.
 PF 23-JUL-1999; 99DE-1034720.

XX 23-JUL-1999; 99DE-1034720.
 XX (TIGR-) TIGR INST GENOMIC RES.
 PA (QUTA-) QUITAGEN GMBH.
 PA (GBFE) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 XX WPI; 2001-160596/17.
 XX New DNA encoding iron or heavy metal transporters, useful for promoting growth of plants on polluted soil -
 PT Claim 3a; Page 21-22; 54pp; German.
 XX This invention describes novel DNA sequences (I) that encode iron or heavy metal transporters (II). The invention also describes (1) a recombinant expression vector (III) containing (I); (2) prokaryotic or eukaryotic cells (III) transformed with (I) or (III); (3) production of (II) by culturing (III); (4) (partial) expression products (IV) of (I) and synthetic proteins or peptides with the same sequences; (5) antibodies (Ab) specific for (IV); (6) hybridoma cells that produce monoclonal Ab; and (7) transgenic plants that contain (III). The iron or heavy metal transporters encoded by (I) transport heavy metal ions across cell walls and regulate homeostasis of trace elements. (I), and their fragments are useful for: (1) expression of (II); (2) as probes and primers for detection, isolation and amplification of full length cDNA sequences; and (3) producing transgenic plants. (III) are used to promote growth, development and yield of plants, particularly leguminosae, especially when growing in soil polluted by heavy metals injurious to plants. They also improve homeostasis of iron and trace elements. Host cells that express the iron or heavy metal transporters take up heavy metals, so reduce pollution of soil and release iron and other trace metals which improves soil quality and protect plants against pollutants from the soil.
 XX Sequence 1173 BP; 236 A; 378 C; 362 G; 196 T; 1 other;
 SQ

Alignment Scores:

Pred. No.: 1-01e-14 Length: 1173
 Score: 271.50 Matches: 96
 Percent Similarity: 42.15% Conservatives: 49
 Best Local Similarity: 27.91% Mismatches: 148
 Query Match: 6.73% Indels: 51
 DB: 22 Gaps: 12

US-09-936-377-2 (1-758) x AAF26431 (1-1173)

QY 430 SerPheGlyValGluGlnAlaAsnTrp-----AspAsnPheThrLeuGluGly 446
 Db 85 GCGCTGTTTCATGCTGGAGCAG-----TGGCAAGCCACCGACCGCTGAACCTGAGCCTG 138
 QY 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspArg 466
 Db 139 GCGCGCGCATGGAG-----CACCCCGGTAGACCCGACGCCCAAGGCAAC 186
 QY 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSer 486
 Db 187 GAGAACTTCGTCGAC-----GGCAGACGCGCCAGCTTCAACGCC 228
 QY 487 PheAlaLeuSerGlyAsn-----TrpTyrPheThrProGlnHisLysLeuSerLeuThr 504
 Db 229 TTCAGCTGTGTCGGGGCTGTGTACCAAGCTCGACCAATCTGTGTCGCTGCGCGCAAC 288
 QY 505 AlaSerHisGlnArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisVal 524
 Db 289 CTCGGCTACACGACGCGGCCCTACCTTTTACGAGCTGTACGCCAATGCGCGCACGTA 348
 QY 525 AlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnIle 544
 Db 349 GCCACCGCGCGGCTTGAAGTCGGTGACCGCCCTGACACAGGAAAGGCCATTTTCGCC 408


```

QY 184 AsnLeuGluLys-----LeuThrSer 190
Db 292200 TTATCGAAAAGGATGTTTCTGTAATTGGCGCAAAATGCAAAATTCGATGTTCTTATTAGT 292259
QY 191 Gly-----GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGlu 206
Db 292260 GGTTCCTAATAATCGGATATAATTACCGACTGGTAAGGCAACAAAGCTAATAATAATACC 292319
QY 207 GlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg-----Asn 222
Db 292320 GCC---ATAAACAGTTTGGGGCTTAGCAAAATTCGGTTGGCAAAATTAATGATGCGAAC 292376
QY 223 LeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeu 242
Db 292377 CCGGTGGAAATTATCCACCGCAAACTCGTTTAA--ACAACAACAGCACCAAGCAATAATG 292434
QY 243 GlyTyrArgLys-----ArgPheTyrArgArgTyrTyrSerAsp 255
Db 292435 AGTGGAAAACGAACTTACTAATGAACAAATTAACAGATCAAAATCAAAAAGTTCACCGAC 292494
QY 256 ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAsp----- 271
Db 292495 AAAAAAGCATCTCTCTCTCTCTACCAACACACCATCACCATCAGAAAGATCAGAGTTT 292554
QY 272 -----CysHisAlaAspIleIle-----TrpGlnLysSerLeu 282
Db 292555 ACTCTAAAGTGAACACAGCTTTAGGTAGTGTCTAGTTATTATTAAGTGCATCAAAATTCCTG 292614
QY 283 IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyr 302
Db 292615 ATCAAGACAGGTATTTAAC-TATTAT-----TTAAGCCA----- 292649
QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 322
Db 292650 GATAATCCTTATCTA-----AATACGCATTCGCAGCTGAT 292685
QY 323 AsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342
Db 292686 ACAATAAATCTATTGAGAAGAACAGCGTAAGTCGTGCTGTAAGATCAAGATCAGACTAAA 292745
QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
Db 292746 TTGACTACCGAGGT-----ATAAATTAGTAATCTTCCGAAATATCT 292790
QY 359 -----AspTyrHisHisAspGluLysAlaGlyAspAlaVal 370
Db 292791 CACATTTCTCTTTTATGGGGTGGATATATCGAGATAAAATCCGTACCGAA----- 292844
QY 371 GluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle 390
Db 292845 -----CGAGGCACAAACGGTAGCGTGCAGTTTCGAGCGGACCCCTAT 292889
QY 391 GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerAlaLeuSer 410
Db 292890 AATGCGAATTCAAACACTACAGCGGTT---TATTTA----- 292922
QY 411 AlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer 430
Db 292923 -----ATCGCCCATATTCCG 292937
QY 431 PhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgVal 450
Db 292938 CTATTTGGG----- 292946
QY 451 GluLys-----GlnLysAlaSerIleArgTyrAsp-----LysAla 462
Db 292947 GAAAAATTTGCTAGTTTCGCCAAGTGTAGCTTATGACCACCTACGTATCCTCAAGTAAACC 293006
QY 463 LeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln 482
Db 293007 GTAAAAATACAAGGATAATCATTTA----- 293030

```

```

QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSer 502
Db 293031 -----TCCTCTCCCAAAATAAATCTGGATAGTAGACC-----AATTGGTATGAT 293075
QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
Db 293076 TTTACTGCCAATATATATGAAGCTTTCCGAGCACCATCTATGCAAGAGCGATTTGTGAGT 293135
QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
Db 293136 GGTGCTCACTTTGGGCAAAATACTCTAGGCTAGATCACATCAATAGATTGTAGCAAAAT 293195
QY 534 LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeu----- 546
Db 293196 CCAAAATTTGGCCCTCAACAGCGAAAAATAAAGAAATACCCGCAAAATCTACATTTTGTAT 293255
QY 547 AlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
Db 293256 AGTCTGTTTAAACAAGCGGTAAATTCAAAATTTGAAGCGACTTATTTCGTAATGATGTG 293315
QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
Db 293316 AAAGATTTTATTAACTTAAATAATTTAATGATGCAAGACAAGTGCAGTCAGGTGCA 293375
QY 580 GlyProLysSer-----IleGluAspAspSerGluMetLysLeuValArgTyr 595
Db 293376 AATCCAAATACAAATGAGCGATTTGTCCAAAAAATTCACAGTATCAA----- 293423
QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
Db 293424 AACATTAATTAATGCCGTTTAAAGCGTATTGAATTCGAAGTCAATACCAA---ACAGAA 293480
QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 293481 CGTTTAACGCTA-----TTTACTAATATGGCAGC 293510
QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 293511 ACCAAAGGTAAGATAAGATAGTGGC-----GAAAGCT 293543
QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
Db 293544 TTATCAACATTCGCCGACGCAAAATCGGGTAGGGTAAATTTATGCTTTAGTAAAGAC 293603
QY 675 -----ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
Db 293604 AAATTCACGTTGGGAGCGACAGTAACCCATTACGCT-----GCTCAACGCGAGTGCCT 293657
QY 693 ArgTyrGluThrArgThr---ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
Db 293658 AAAGATCATAGTGTATACCTACCCCAAGTTATATATAGTACCGACCTTCGTCTACCTATGCA 293717
QY 712 ArgAsnThrArgTyrGlyGluTyr 719
Db 293718 -----CCATTAAGGCGAATCG 293735
RESULT 26
AAS97293
ID AAS97293 standard; DNA; 2145 BP.
XX
AC AAS97293;
XX
DT 12-MAR-2002 (first entry)
XX
DE Neisseria meningitidis virulence gene #98.
XX
KW Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
XX infection; Gram-negative bacteria; antimicrobial; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200185772-A2.
XX

```



```

QY 460 pLysalaleuileAspArgGluAsnTyrTyrlsGlnProLeuProAspLeuGlyAlaHi 480
Db 1332 CGGCAAAACCGTTTCAAGCAACACCTTAAC-----CCAGTTTCGGCGTGAT 1379
QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisly 500
Db 1380 TTGGCAGCGCACGAA-----CACTGGAGTTCAGCGCAGCCACAA 1421
QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db 1422 CTAC-----GCCAGCGCAGCGCGCGCTG-----TATGAGCGCGCTGCAAAACCA 1466
QY 520 sGlyLyshis-----ValAlaThrAsnThrPheGluValGlyAsnLyshis 536
Db 1467 CGCAAAACCGGCATCATCTCGATTGCCAGCGCAG----- 1503
QY 536 uAsnLySgluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGl 556
Db 1504 -AAAGCGGAACGCGCGCAATACCGAAATCGCTTCAACTACACGACGCGCAGTTTCG 1562
QY 556 nTyrAsn-----LeuAlaLeuTyrArgAsnAr 565
Db 1563 CGCAAAACCGCAGCTACTTCTGCAGACCATCAAGACGCGCTTGCCAATCCGCAAAACCG 1622
QY 565 gPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
Db 1623 CCACAGCTCTGCGCGTTCGGTGAAGCGTCAATGCCGTTACATCAAAACACACGGTTA 1682
QY 579 -----ArgGl 580
Db 1683 CGAATTGGCGCGTCTCTACCGCAGCGCGCTGACTGCGCAAGTCGGGTGAAGCCACAG 1742
QY 580 yProLysSerIleGluAspSerGluMetLySleuValArgTyrAsnGlnSerGlyAl 600
Db 1743 CAACCGCGCTTTTACGATACGCAACAGCAAGCTGTTGAGCGCGAATCTCGAA----- 1797
QY 600 aAspPheTyrGlyValaGlu-----GlyGluIleTyrPhelysPr 613
Db 1798 -----TTTGGGCACAAAGTCGCGCACCTTGGACGGTTCCTTGCCTACCGTTCGA 1850
QY 613 oThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLyAsnLe 633
Db 1851 AAACCCGAATCTGGAATCGGTCGCGCGCGCTATGTT----- 1890
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAs 653
Db 1891 -----CAAAAAGCCGCTGGTTCGATATTGGTGGCAGGTCAAAAAGA 1931
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLySAlaSerLeuTh 673
Db 1932 CGCAACGCGCAATTTGAAAACGTTGTAGCAAGGTTTCGGTGTGAAC----- 1980
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLySleuAlaAr 693
Db 1981 -----GATGTCCTTCGCCCACTGGAAACCGCTGGG 2009
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
Db 2010 CAAGACACGCTCAAT-----GTTAATCTTCGGTTAAC----- 2043
QY 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLySAlaAspAsnLeuLeuAsnGlnSerVa 733
Db 2044 -----AACGTGTTCAACACGTTCTA 2063
QY 733 lTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThrGl 752
Db 2064 CTATCGCACACCAACGATGACCAATACCTCGCGCGCGTGGGACGTGATGTAGCTTT 2123
QY 752 yGlyValAsnValLySPh 758
Db 2124 GGGCGTGAACACTCAAGTTT 2142

```

```

RESULT 27
AAA81477/c
ID AAA81477 standard; DNA; 26778 BP.
XX
AC AAA81477;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_25 SEQ ID NO:25.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
XX used in the diagnosis and treatment of N. meningitidis infection and
XX other Neisserial infections, for example, N.gonorrhoea -
XX
XX Claim 7; Page 524-531; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414
XX represent specifically claimed Neisseria meningitidis genomic DNA
XX sequences; AAA81360 to AAA81303 and AAB25620 to AAB25663 represent
XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
XX isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
XX sequences, which are all used in the exemplification of the present
XX invention. The nucleic acid sequences, protein sequences, and antibodies
XX against them, can be used in the manufacture of a composition. The
XX composition can be used as a medicament (or in the manufacture of a
XX medicament) for treating, preventing or diagnosing infection due to
XX Neisserial bacteria. For example, some of the identified proteins could
XX be components of vaccines against Meningococcus B; against all serotypes;
XX and/or against all pathogenic Neisseriae. Identification of sequences
XX from the bacterium will also facilitate production of biological probes,
XX particularly organism-specific probes. Attempts to make efficacious
XX Meningococcus B vaccines have failed mainly due to antigen tolerance.
XX Multivalent vaccines have also been tried but none have successfully
XX overcome antigenic variability. The provision of further complete
XX sequences may provide an opportunity to identify secreted or surface
XX exposed proteins that may be presumed targets for the immune system and
XX which are not antigenically variable or at least more conserved than
XX other more variable regions.
XX
XX Sequence 26778 BP; 6602 A; 7008 C; 6846 G; 6318 T; 4 other;

```

```

Alignment Scores:
Pred. No.: 4.48e-09 Length: 26778
Score: 229.50 Matches: 197
Percent Similarity: 33.45% Conservative: 93
Best Local Similarity: 22.72% Mismatches: 301
Query Match: 5.69% Indels: 277

```

DB: 21 Gaps: 46
US-09-936-377-2 (1-758) x AAA81477 (1-26778)
QY 16 IleAsnThrProLeu-----LeuAla 22
DB 6425 ATGAATACCCCATTTTCCTCTCAGCGCTCTCGCTTACCCTGGCGGAGGTTTGGC 6366
QY 23 GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer 42
DB 6365 CATCGCGCAGAAATAATGCAAGGTCTGATACCGTATCCGTAAAAAGCGACCGC 6306
QY 43 ArgProArgAlaThrSerGlyLeuLeuHisThrThrAlaSerAspLysIleLeuSer 62
DB 6305 CAA-----GGCAGCAAAATCCGTACCAACATCGTT--- 6276
QY 63 GlyAspThrLeuArgGlnLys-----AlaValAsnLeuGlyAspAlaLeuAsp 78
DB 6275 -----ACGCTGCAACAAGAGCAAGAACCGCAACCGATATGCGCGAACTTTAAAA 6222
QY 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIle--- 97
DB 6221 GAAGAGCCCTCATC-----GATTCGGCGGCGCAACGCGCACGTCCCAATTCCTG 6171
QY 98 -----ArgGlyGlnThrGlyArgArgIleLysVal----- 107
DB 6170 ACCTGCGCGCATGGTCAAACTCTGTGCACATCAAGGTGACACGCGCTATTCCGAC 6111
QY 108 -----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
DB 6110 AGCCAATCTTTACCACCAAGCAGA----- 6084
QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
DB 6083 -----TTTATTGTGATCCCGCTTTGGTTAAAGTGGTTCCGTACAAAAGGC----- 6036
QY 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
DB 6035 -----CGCGTTCCGCTCTCGCGGTATCGCGCGACCAACGCGCGCATATC 5988
QY 165 GluLysMetProLys-----AsnGlyValSerGlyGluLeuGlyLeuArg 179
DB 5987 ACCAAACCGTCGATCCCAAGACCTGCTCAAGGCTTGGATAAAACCTGGCGCGTGC 5928
QY 180 LeuSerSerGly-----AsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeu 197
DB 5927 CTCACAGCGGCTTCCGACCAAGAGCGGTAGCTACGCGCAAGCGTATTCGGGAA 5868
QY 198 GlyLysAsnPheValLeuHisThrGluGlyLeuTyr-----ArgLysSerGly 213
DB 5867 GAGGGCAACTTC-----GACGCGTTGTTCTTACAAACCGCAACATGAANA 5820
QY 214 AspTyrAlaValProArg---TyrArgAsnLeu-----LysArgLeuProAsp 228
DB 5819 GATTACGAACAGGTAAAGCTTCCGTAATAATTCAACGGCGGCAAAACCTGATCCGTAC 5760
QY 229 SerPro-----ArgArgPheAlaAsnGlyGln 237
DB 5759 AGCGCGCTGCAACACGACGTACCTCGCCAAATCGGAACAGCTTCGCGACGCGAC 5700
QY 238 HisArgAlaValLeuGlyTrpArgLysArgPheTyrArgThrTyrSerAspArgArg 257
DB 5699 CACCGCATCGTATTG-----AGCCATATGA 5673
QY 258 AspGlnTyr---GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIle 276
DB 5672 GACCAGACCGGGCATCTGACCGCTCGTGAAGAAATTT-----ACCGTC 5628
QY 277 IleTrpGlnLysSerLeuIleAsn---LysArgTyrLeuGlnLeuTyrProHisLeuLeu 295
DB 5627 GCGCGCGATAAAGAGCGAATAGTATGGAACGCCACCGCCCTGCTTACCGCGAA---ACC 5571
QY 296 ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp 315

DB 5570 ACACAATCCAAACACCAATTTGGCGGTACACGGTAAAAAACCTGGCTTTGTGAAAAATCG 5511
QY 316 AspAlaHisAlaHisAlaHisAsnGlyLysProTyrPheAspLeuArgAsnLysArgTyr 335
DB 5510 GATGCCAACCCCTTATGTGTT---GGAAAAAGAACGCTATTCCGCGCATGACGCGCAC 5455
QY 336 GluLeuArgAlaGluTrpLys-----GlnProPhePro-----GlyPheGlu 349
DB 5454 CGGTTACGAGCAATGTAAAAAGGCCCAACCATACCCAAATCACCACTCGGGGTATGAA 5395
QY 350 ---AlaLeuArg-ValHisLeuAsnArgAsn-----AspTyrHis 361
DB 5394 CTTCAACTTCGACAGCGCGCTTCGCCGAACAAACCTGCTGAAATACGGTATCAACTACCG 5335
QY 361 sHisAspGluLysAlaGlyAspAlaValAlaGluAsnPhePheAsnAsnGlnThrGlnAsnAl 381
DB 5334 CCAATCAGGAATCAAAACCGCAAGC-----TTTTGAATTCACAA-----TT 5293
QY 381 aArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyr 401
DB 5292 TAAATTTGAAGATAAAAGAAAAAGCAACTGATGAAGAGAAAAAATAAG----- 5247
QY 401 rLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLe 421
DB 5246 -----AACCGTGAATAATGAATAATTCGCAAGCCTACCG 5212
QY 421 uLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsn---TrpAs 440
DB 5211 TCTGACCAACCCGACCAACCATACCGCGGTATATCGAAGCCATTCACGAGATTGA 5152
QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
DB 5151 CGGCTTTACCTGACCGCGGCTGCTTACGACCGCTTCAAGGTGAAA---ACCACGA 5095
QY 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHis 480
DB 5094 CGCAAAACCGTTTCAAGCAACACCTTAAC-----CCGAGTTTCGCGGTAT 5047
QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLys 500
DB 5045 TTGGCAGCGCACGAA-----CACTGAGCTTCAGCGCGACCCACAA 5005
QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis 520
DB 5004 CTAC-----GCCAGCGCAGCGCGCGCTG-----TATGACGCGCTGCAAAACCCA 4960
QY 520 sGlyLysHis-----ValAlaThrAsnThrPheGluValGlyAsnLysHisLe 536
DB 4959 CGCAAAACCGCGCATCTCTGATTCCGACGCGCAG----- 4923
QY 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpG 556
DB 4922 -AAAGCCGAACGCGCGCAATACCGAAATCGCTTCACTACACGACGCGACGTTTCG 4864
QY 556 nTyrAsn-----LeuAlaLeuTyrArgAsnAr 565
DB 4863 CGCAAAACCGCAGCTACTTCTGCGACACCATCAAGACGCGCTTGCATCCGCAAAACCG 4804
QY 565 pPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
DB 4803 CCACGACTCTGTGCGCGTCCGTGAAGCGCTCAATCGCGTTACATCAAAACCCACGTTTA 4744
QY 579 ----- 580
DB 4743 CGAATTGGCGCGTCTTACCGCACCGCGCTGACTGCTCAAGTCGCGGTAGCCACAG 4684
QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
DB 4683 CAAACCGCGCTTTTACGATACGCAAAAGACAGCTGTGAGCGCGAATTCCTGAA----- 4629
QY 600 aAspPheTyrGlyAlaGlu-----GlyGluIleTyrPheLysPr 613

Db 4628 -----TTTGGCGCACAAAGTCGGCGCCTTGGACGCTTCCCTTGCTACCGCTTCCA 4576
 Qy 613 oThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLe 633
 Db 4575 AAACCCGAATCGGAATCGGCTGCGCGCGCGTATATGTT----- 4536
 Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
 Db 4335 -----CHAAAGCCGTGGTTCGATATGTTGGCGAGGTCAAAAGA 4495
 Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
 Db 4494 CCGCAACCGCAATTTGAAAAAGTTGTACGCAAGGTTTCGGTGTGAAC----- 4446
 Qy 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
 Db 4445 -----GATGTCTTCGCCCACTGGAAACCGCTGGG 4417
 Qy 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
 Db 4416 CAAGACACGCTCAAT-----GTTAACTTTTCGGTTAAC----- 4383
 Qy 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerVa 733
 Db 4382 -----AACGGTTCACACAGCTTCTA 4363
 Qy 733 lTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThrGl 752
 Db 4362 CTATCCGACAGCCACCAATGATGACCAATACCCCTGCCGGCGTGGACGTGATGACGCTT 4303
 Qy 752 yGlyValAsnValLysPhe 758
 Db 4302 GGGCGTGAACACAAAGTTTC 4284

RESULT 28

AAF21612
 ID AAF21612 standard; DNA; 349980 BP.
 XX
 AC AAF21612;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200066791-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05928.
 XX
 PR 30-APR-1999; 99US-0132068.
 PR 08-OCT-1999; 99WO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
 PI Frazer CM, Grandi G;
 XX
 DR WPI; 2000-647603/62.
 XX
 XX Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections -
 XX
 PS Claim 7; Appendix A; 692pp; English.

XX The present invention describes the full length genome of
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.
 XX
 SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

Alignment Scores:
 Pred. No.: 1.55e-07 Length: 349980
 Score: 229.50 Matches: 197
 Percent Similarity: 33.45% Conservative: 93
 Best Local Similarity: 22.72% Mismatches: 301
 Query Match: 5.69% Indels: 277
 DB: 21 Gaps: 46

US-09-936-377-2 (1-758) x AAF21612 (1-349980)

Qy 16 IleAsnThrProLeu-----LeuAla 22
 Db 292611 ATGAATACCCCAATGTTCCGCTCTCAGCTCTGCTTACCTGCGCGCAGGTTTGGC 292670
 Qy 23 GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer 42
 Db 292671 CATCGGCAGAAATAATGCAAGTGTCTGATACCGTTACCGTAAAGGCGACCGC 292730
 Qy 43 ArgProArgAlaThrSerSerGlyLeuHisThrThrAlaSerAspLysIleIleSer 62
 Db 292731 CAA-----GGCAGCAAAATCCGTACCAACATGTT--- 292760
 Qy 63 GlyAspThrLeuArgGlnLys-----AlaValAsnLeuGlyAspAlaLeuAsp 78
 Db 292761 -----ACGCTGCAACAAAAAGCAGCAACCGCATATGCGGAACTCTTAAAA 292814
 Qy 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIle--- 97
 Db 292815 GAAGAGCCCTCCATC-----GATTCGGCGCGCAACGCGACGATCCCAATTCCTG 292865
 Qy 98 -----ArgGlyGlnThrGlyArgArgIleLysVal----- 107
 Db 292866 ACGCTGCGCGCATGGGTCAAACTCTGACATCAAGTGGACACCGCTATTCCGAC 292925
 Qy 108 -----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 Db 292926 AGCCAAATCTTTTACCACCAAGGAGGAG----- 292952
 Qy 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
 Db 292953 ---TTTATTGTCGATCCCGCTTTGGTTAAAGTGTGTTTCGTACAAAAGGC----- 293000
 Qy 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
 Db 293001 -----GCGGTTCCGCTCTGCGGATATCGCGCGACCAACCGCGCATCATC 293048
 Qy 165 GluLysMetProGlu-----AsnGlyValSerGlyGluLeuGlyLeuArg 179

```

Db 293049 ACCAAACCGTCGATGCGCCACAGACCTGCTCAAGGCTTGGATAAAACCTGGGCGCTGGCG 293108
Qy 180 LeuSerSerGly-----AsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeu 197
Db 293109 CTCACACGGGCTTGGCCAGCAACGAGCGGTAAAGCTAGCGGCGCAAGCTATTCCGGAAA 293168
Qy 198 GlyLysAsnPheValLeuHisThrGluGlyLeuTyr-----ArgLysSerGly 213
Db 293169 GAGGCAACCTTC-----GACGGCTTGTCTCTTACACCGCAACCAATGAAAAA 293216
Qy 214 AspTyrAlaValProArg---TyrArgAsnLeu-----LysArgLeuProAsp 228
Db 293217 GATTACGACGAGGTAAAGCTTCGTTAATAATTCAACGGCGGCAAAACCGTACCGGTAC 293276
Qy 229 SerPro-----ArgArgPheAlaAsnGlyGln 237
Db 293277 AGCGCGCTGGCAACGCGAGCTACCTCGCCAAAATCGGAACAAGCTTCGGCGACGCGAC 293336
Qy 238 HisArgAlaValLeuGlyTrpArgLysArgPheTyrArgThrTyr-SerAspArgArg 257
Db 293337 CACCGCATCGTATTG-----AGCCATATGAAA 293363
Qy 258 AspGlnTyr---GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIle 276
Db 293364 GACCAGCACCGGGCATCGGTACCGTCCGTGAAGATT-----ACCGTC 293408
Qy 277 IleTrpGlnLysSerLeuIleAsn---LysArgTyrLeuGlnLeuTyrProHisLeuLeu 295
Db 293409 GCGCGCGATAAGAGCGAATAAGTATGAAGCCCAAGCCCTCTTACCGGAA---ACC 293465
Qy 296 ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAsp 315
Db 293466 ACACATCCACACCAATTTGGCGTACACGGGTAAACCTGGGCTTTGTGAAAAAATG 293525
Qy 316 AspAlaHisAlaHisAlaHisGlnGlyLysProTrpIleAspLeuArgAsnLysArgTyr 335
Db 293526 GATGCCAACGCGCTATGTGT---GAAAAAGAACGCTATTCCGCGCATGACGCGGCAC 293581
Qy 336 GluLeuArgAlaGluTrpLys-----GlnProPhePro-----GlyPheGlu 349
Db 293582 CGGTTACCGACGCAATGTAAAGAGCGCCCAACCATACCCAAATCACCCTCGGGTATGAA 293641
Qy 350 --AlaLeuArg-ValHisLeuAsnArgAsn-----AspTyrHis 361
Db 293642 CTTCAACTTCACACCGCCCTTGGCGAACACCTGCTGAATACGGTATCAACTACCG 293701
Qy 361 sHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAl 381
Db 293702 CCATCAGGAATCAAAACCGCAAGCG-----TTTTGTAATTCACAA-----TT 293743
Qy 381 aArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyr 401
Db 293744 TAAATTTGAATGAATAAGAAAGAACACTGATGACAGAGAAAATAAG----- 293789
Qy 401 rLeuGlyGlnLysSerSerAlaLeuSerAlaThr-SerGluAlaValLysGlnProMetLe 421
Db 293790 -----AACCGTGAATAATGAAAAATGCAAAATGCAAAAGCTACCG 293824
Qy 421 uLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn---TrpAs 440
Db 293825 TCTGACCAACCCCGCAACCAACCGATACCGGCGCTATATCGAAGCATTCACGAGATTGA 293884
Qy 440 pAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db 293885 CGGCTTTACCTGACCGCGCGGCTGCGTTACGCGCTTCAAGGTGAAA---ACCCACGA 293941
Qy 460 pLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyValAla 480
Db 293942 CGGCAAAACCGTTTCAAGCAACAACCTTAAC-----CCGAGTTTCGCGGTGAT 293989
Qy 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLy 500

```

```

Db 293990 TTGGCAGCGGCACGAA-----CACTGGAGCTTTCAGCGCGAGCCACAA 294031
Qy 500 sLeuSerLeuThrAlaSerHisGlnArgLeuProSerThrGlnGluLeuTyrAlaHis 520
Db 294032 CTAC-----GCCACCGCAGCCCGGCGCTG-----TATGACGGCGCTCAAAACCA 294076
Qy 520 sGlyLysHis-----ValAlaThrAsnThrPheGluValGlyAsnLysHisLe 536
Db 294077 CGCAAAACCGGCATCATCTCGATTGCGGACGCGCAG----- 294113
Qy 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpG 556
Db 294114 -AAAGCCGAACCGCGCGCAATACGAAATCGGCTTCAACTACACGACGCGACCTTTGC 294172
Qy 556 nTyrAsn-----LeuAlaLeuTyrArgAsnArg 565
Db 294173 CGCAAAACCGCAGCTACTTCTGGCAGACCATCAAGACCGGCTTGCCTATCCGCAAAACCG 294232
Qy 565 pPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
Db 294233 CCACGACTCTGTCGCGCTCGTGAAGCGCTCAATGCGGTTTACATCAAAACCAACCGTTA 294292
Qy 579 -----ArgGln 580
Db 294293 CGAATGGCGCGCTCTTACCGCAGCGCGGCTGACTGCCAAAAGTCGCGTAAGCCACAG 294352
Qy 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db 294353 CAAACCGCGCTTTTACGATACGCACAAAGACAGCTGTGTAGCGGCAATCTCGAA----- 294407
Qy 600 aAspPheTyrGlyAlaGlu-----GlyGluIleTyrPheLysPr 613
Db 294408 -----TTTGGCGCACAAAGTCGCGCGCAGCTTGGACGGCTTCCCTTGCCTACCGCTTCCA 294460
Qy 613 oThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLe 633
Db 294461 AAACCCGNAATCGGAATCGGCTCGCGCGCGCTTATGTT----- 294500
Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 294501 -----CAAAAAGCGCTGGTTCGNATATTGGTGGCAGGTCAAAAAGA 294541
Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 294542 CCGCAACGCGCAAAATGGAACAGTTTACGCAAAAGGTTTCGGTGTGAAC----- 294590
Qy 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 294591 -----GATGCTCTCCCAACTGGAAACCGCTGGG 294619
Qy 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 294620 CAAGACACGCTCAAT-----GTTAATCTTTCCGTTTAA----- 294653
Qy 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerVa 733
Db 294654 -----AACGTGTTCACACGCTTCTA 294673
Qy 733 lTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThrGl 752
Db 294674 CTATCCGCACAGCCCAACGATGGACCAATACCTCCCGGGCGTGGACGCTGATGTACGCTT 294733
Qy 752 yGlyValAsnValLysPhe 758
Db 294734 GGGCGTGAACCTACAAGTTTC 294752

```

RESULT 29
 ID AAA81489
 XX AAA81489 standard; DNA; 837096 BP.
 AC AAA81489;
 XX
 DT 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.
 DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.
 XX Neisseria meningitidis.
 OS Neisseria meningitidis.
 PN WO200022430-A2.
 PD 20-APR-2000.
 XX 08-OCT-1999; 99WO-US23573.
 XX 09-OCT-1999; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX (CHIR) CHIRON CORP.
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX WPI; 2000-318079/27.
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -
 XX Claim 7; Page 629-865; 1760pp; English.
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;
 SQ
 Alignment Scores:
 Pred. No.: 5,166-07 Length: 837096
 Score: 229.50 Matches: 197
 Percent Similarity: 33.45% Conservative: 93
 Best Local Similarity: 22.72% Mismatches: 301
 Query Match: 5.69% Indels: 277
 DB: 21 Gaps: 46
 US-09-936-377-2 (1-758) x AAA81489 (1-837096)
 QY 16 lleaThrProLeu-----LeuAla 22
 DB 657382 ATGAATACCCCATTTGTCGTCACGCTCTCGCTTACCTCGCGCGAGTTTTCGCC 657441

QY 23 GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer 42
 DB 657442 CATCGCGCAGAAATAATGCAAGGTCTGATCGTATACCGTAAAGGCGACCGC 657501
 QY 43 ArgProArgAlaThrSerGlyLeuHisThrSerThrAlaSerAspLysIleLeuSer 62
 DB 657502 CAA-----GGCAGCAAAATCCGTACCAACATCGTT--- 657531
 QY 63 GlyAspThrLeuArgGlnLys-----AlaValAsnLeuGlyAspAlaLeuAsp 78
 DB 657532 -----ACGCTGCAACAAAAAGACGAAAGACCGCAACCGATATGCGCGAATCTTTAAAA 657585
 QY 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIle--- 97
 DB 657586 GAAGAGCCCTCATC-----GATTTCCGCGCGCGCAACGCGACGCTCCCAATTCTG 657636
 QY 98 -----ArgGlyGlnThrGlyArgArgIleLysVal----- 107
 DB 657637 ACGCTGCGCGGATCGGTCAAACTCTGTCACATCAAGTGGACCAACGCCCTATTCCGAC 657696
 QY 108 -----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 DB 657697 AGCCAAATCTTTTACCACCAAGGCAGA----- 657723
 QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
 DB 657724 ---TTTATTGTCGATCCCGCTTTGTTAAAGTCGTTTCGTCAAAAAGGC----- 657771
 QY 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
 DB 657772 -----GCGGTTCCGCTCTCGCGGTATCGCGCGCACCAACGCGCGATCATC 657819
 QY 165 GluLysMetProGlu-----AsnGlyValSerGlyGluLeuGlyLeuArg 179
 DB 657820 ACCAAACCGTCGATGCCACCAACCTGCTCAAGGCTTGATGATAAAACGCGGCGC 657879
 QY 180 LeuSerSerGly-----AsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeu 197
 DB 657880 CTCACACGCGGCTTTGCCAGCAACGAGCGTAAAGCTACGCGCGCAAGCTATTTCGGGAAA 657939
 QY 198 GlyLysAsnPheValLeuHisThrGluGlyLeuTyr-----ArgLysSerGly 213
 DB 657940 GAGGCAACTTC-----GACGCTTGTCTCTTACACCGCAACCAATGAAAAA 657987
 QY 214 AspTyrAlaValProArg---TyrArgAsnLeu-----LysArgLeuProAsp 228
 DB 657988 GATTACGAAGCAGGTAAAGGCTTCGTAATAATTTCAACGCGCGCAAAACCGTACGTAC 658047
 QY 229 SerPro-----ArgArgPheAlaAsnGlyGln 237
 DB 658048 AGCGCGTGGACAAACGCGAGCTACCTGCCAAATCGGAACAGCTTCGCGCGACGCGAC 658107
 QY 238 HisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArg 257
 DB 658108 CACCGCATCGTATTC-----AGCCATATGAAA 658134
 QY 258 AspGlnTyr---GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIle 276
 DB 658135 GACCAGCACCGGGCATCCGTACCGTCCGTGAAGATT-----ACCGTC 658179
 QY 277 IleTyrGlnLysSerLeuIleAsn---LysArgTyrLeuGlnLeuTyrProHisLeuLeu 295
 DB 658180 GCGCGCGATTAAGACGCAATTAAGTATGGAAGCCCAAGCCCTCTGTTACCGCAA---ACC 658236
 QY 296 ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAsp 315
 DB 658237 ACACATCAACACCAATTTGGCGTACAGGGTAAAAAAGCTGGGCTTTGTTCGAAAAAATG 658296
 QY 316 AspAlaHisAlaHisAlaHisAsnGlyLysProTyrPilleAspLeuArgAsnLysArgTyr 335
 DB 658297 GATGCCAACGCTTATGTGT---GGAAAGAACGCTATTTCGCGCGATGACAGCGGCAC 658352

```

QY 336 GluLeuArgAlaGluTrpLys-----GlnProPhePro-----GlyPheGlu 349
Db 658353 CGGTTACGAGCAATGTAAGAGCGCCCAACCATACCAATCACCACGCGGTATGAA 658412
QY 350 ---AlaLeuArg-ValHisLeuAsnArgAsn-----AspTyrHis 361
Db 658413 CTTCAACTTCGACGCGGCTTCGCGAACAACCCCTGCTGAAATACGGTATCAACTACCG 658472
QY 361 sHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAl 381
Db 658473 CCATCAGGAATCAAAACCGCAAGCG-----TTTTTGAATTCACAA-----TT 658514
QY 381 aArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTy 401
Db 658515 TAAATTTGAAGATAAAGAAAGAACCAACTGATGAGAGAGAAATAAG----- 658560
QY 401 rLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLe 421
Db 658561 -----AACCGTGAATGMAAAATGCGCAAGCCTACCG 658595
QY 421 uLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn---TrpAs 440
Db 658596 TCTGACCAACCCCAACCAACCATACCGCGGGGTATATCGAAGCCATTCACGAGATTGA 658655
QY 440 pAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAs 460
Db 658656 CGGCTTTACCTGACCGCGGGGTGCGTTACGACCGCTTCAAGGTGAAA---ACCCACGA 658712
QY 460 pLysAlaLeuLeuAspArgGluAsnTyrTyrLysGlnProLeuPheAspLeuGlyAlaHi 480
Db 658713 CGGCAAAACCGTTTCAAGCAACACCTTAAC-----CCGAGTTTTCGGGTGAT 658760
QY 480 sArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLy 500
Db 658761 TTGCGCAGCCGACGAA-----CACTGAGCTTCAGCGCGAGCCACAA 658802
QY 500 sLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db 658803 CTAC-----GCCAGCGCAGCCGCGCTG-----TATGACGCGCTGCAACCCCA 658847
QY 520 sGlyLysHis-----ValAlaThrAsnThrPheGluValGlyAsnLysHisLe 536
Db 658848 CGGCAAAACCGGCATCATCTCGATTGCGCAGCGCAGC----- 658884
QY 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGl 556
Db 658885 -AAAGCCGAACGCGCGCAATACCGAAATCGGCTTCACTACAACGACGCGCATTTGC 658943
QY 556 nTyrAsn-----LeuAlaLeuTyrArgAsnAr 565
Db 658944 CGCAAAACGGCAGCTACTTCTGGCAGACCATCAAGACGCGCTTGCCTCCCAATCCGCAAAACCG 659003
QY 565 gPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
Db 659004 CCACGACTCTGTCCGCTCGTGAAGCGCGTCAATGCGGTTACATCAAAACCAACCGTTA 659063
QY 579 -----ArgGl 580
Db 659064 CGAATGGGGCGCTCTACCGCACCGCGCGCGCTGACTGCGCAAGTCGGCGTAAGCCACAG 659123
QY 580 yProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAl 600
Db 659124 CAAACCGCGCTTTTACGATACGACACAAGACAGCTGTTGAGCGCGAATCTCTGAA----- 659178
QY 600 aAspPheTyrGlyAlaGlu-----GlyGluIleTyrPheLysPr 613
Db 659179 -----TTTGGGCGACACAGCTGCGCGCACTTGACGCGCTTCCCTTACCGCTTCCA 659231
QY 613 oThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLe 633
Db 659232 AAACCCGAATCTGGAATCTGCGTGGCGCGCGCTTATGTT----- 659271
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653

```

```

Db 659272 -----CAAAAAGCCGTTGGGTTCGATATTGGTGGCAGGTCAAAAAGA 659312
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 659313 CGCACAACGCAATTCGAAAACGTTGTACGCAAAAGGTTTCGGTGTGAAC----- 659361
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 659362 -----GATGCTTCGCAACTGGAAACCGCTGGG 659390
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 659391 CAAAGACACGCTCAAT-----GTTAATCTTTCGGTTAAC----- 659424
QY 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerVa 733
Db 659425 -----AACGTGTTCAACAGTTCTA 659444
QY 733 lTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThrGl 752
Db 659445 CTATCGCACAGCAACGATGGACCAATACCTCGCGGCGTGGACGTGATGTACGCTT 659504
QY 752 yGlyValAsnValLysPhe 758
Db 659505 GCGCGTGAACATACAGTTTC 659523

```

RESULT 30
 ABQ78300
 ID ABQ78300 standard; DNA; 2307 BP.
 XX ABQ78300;
 DT 05-NOV-2002 (first entry)
 XX
 DE Nucleotide sequence of p64 polypeptide.
 XX
 KW p177; p88; p64; p55; p46; vaccine; gonorrhea; gene; ss.
 XX
 OS Neisseria gonorrhoea.
 FH Key Location/Qualifiers
 CDS 99..2261
 FT /*tag= a
 XX /product= "p64"
 PN WC200260936-A2.
 XX
 PD 08-AUG-2002.
 XX
 PF 31-JAN-2002; 2002WO-US02881.
 XX
 PR 31-JAN-2001; 2001US-266070P.
 PR 06-AUG-2001; 2001US-310356P.
 PR 23-OCT-2001; 2001US-344452P.
 XX
 PA (IOWA) UNIV IOWA RES FOUND.
 PA (REGC) UNIV CALIFORNIA.
 PA (APIC/) APICELLA M A.
 PA (EDWA/) EDWARDS J L.
 PA (GIBS/) GIBSON B W.
 PA (SCHE/) SCHEFFLER K.
 PA (BROW/) BROWN E.
 XX
 PI Apicella MA, Edwards JL, Gibson BW, Scheffler K, Brown E;
 XX
 DR WPI; 2002-619227/66.
 DR P-PSDB; ABB78069.
 XX
 PT New polypeptide comprising p177, p88, p64, p55 or p46 from Neisseria
 PT gonorrhoea, useful for preventing, or protecting a female patient
 PT against, N. gonorrhoea colonization or infection -
 XX

PS Claim 14; Page 127-128; 130pp; English.

XX The present sequence encodes a p64 polypeptide. The specification
 CC describes p177, p88, p55 and p46 polypeptides from Neisseria
 CC gonorrhea. The polypeptides are useful as vaccines, for preventing,
 CC or protecting a female patient against, N. gonorrhea colonization or
 CC infection. Such immunisation can prevent gonorrhea in women.

XX
 SQ Sequence 2307 BP; 656 A; 673 C; 541 G; 437 T; 0 other;

Alignment Scores:
 Pred. No.: 2.3e-10 Length: 2307
 Score: 227.50 Matches: 189
 Percent Similarity: 33.14% Conserv: 101
 Best local Similarity: 21.60% Mismatches: 299
 Query Match: 5.64% Indels: 287
 DB: 24 Gaps: 43

US-09-936-377-2 (1-758) x ABQ78300 (1-2307)

```

QY      16 lleAnthrProLeu-----LeuAla 22
      ::::::::::::::|
Db      99 ATGAATACCCCAATGTCGCTCTAGCCTGCTCTCGCTCACACTCGCGCAGGTTTGGC 158
      ::::::::::::::|
QY      23 GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer 42
      ::::::::::::::|
Db      159 CACGCGCGCAGAAATAATGCCAAGGTCGTGATACCGTTACTGTATAAGCGCGACCGC 218
      ::::::::::::::|
QY      43 ArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysllelleSer 62
      ::::::::::::::|
Db      219 CAA-----GGCAGAAAATCCGTACCAACATTCGT--- 248
      ::::::::::::::|
QY      63 GlyAspThrLeuArgGlnLys-----AlaValAsnLeuGlyAspAlaLeuAsp 78
      ::::::::::::::|
Db      249 -----ACGCTGCACAAAAAGACGAAAGCACCGCAACCGATATGCGCGACTTTAAA 302
      ::::::::::::::|
QY      79 GlyValProGlylleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVallle--- 97
      ::::::::::::::|
Db      303 GAAGAGCGTCGCATC-----GATTTGCGCGCGCGCAACGGCACGTCCTCAATTCCTG 353
      ::::::::::::::|
QY      98 -----ArgGlyGlnThrGlyArgAlleyVal----- 107
      ::::::::::::::|
Db      354 ACGCTGCGCGCGCATGGTGCAGAACTCTGTGCACATCAAGTGGACCAACGCCTATTCGCAC 413
      ::::::::::::::|
QY      108 -----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
      ::::::::::::::|
Db      414 AGCCAAATCCTTTACCACCAAGGCAG----- 440
      ::::::::::::::|
QY      125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleuArgGlyProValThr 144
      ::::::::::::::|
Db      441 ---TTTATGTGCATCCCGCTTTGGTTAAAGTCGTTTCGTACAAAAGGC----- 488
      ::::::::::::::|
QY      145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysllePro 164
      ::::::::::::::|
Db      489 -----GCGGGTTCGCGCTCTGCGGGTATCGCGCGCACCAACGGCGGATCATC 536
      ::::::::::::::|
QY      165 GluLysMetProGlu-----AsnGlyValSerGlyGluLeuGlyLeuArg 179
      ::::::::::::::|
Db      537 GCCAAAACCGTCGATGCCCAAGACCTGCTCAAGAGGCTTGGATAAAAACTGGCGGTCGC 596
      ::::::::::::::|
QY      180 LeuSerSerGly-----AsnLeuGluLysLeuThrSerGlyGlylleAsnIleGlyLeu 197
      ::::::::::::::|
Db      597 CTCACACAGCGGTTTGCACAGACGAGGCGTAAGTACGGCGCAAGCGTATTCGGA 656
      ::::::::::::::|
QY      198 GlyLysAsnPheValLeuHisThrGluGlyLeuTyr-----ArgLysSerGly 213
      ::::::::::::::|
Db      657 GAGGGCAACTTC-----GACGGCTGTCTCTTACAACCGCAACGATGAAAA 704
      ::::::::::::::|
QY      214 AspTyrAlaValProArg---TyrArgAsnLeu-----LysArgLeuProAspSer 229
      ::::::::::::::|
Db      705 GATTACGAGCGCGCAAGGTTTCCCAATGTCAACGGCGGCAAAACCGTACCGTCAGC 764
      ::::::::::::::|
QY      230 Pro-----ArgArgPheAlaAsnGlyGlnHis 238
      ::::::::::::::|

```


QY	407	rAlaLeuSerAlaThrSerGluAlaVallysGlnProMetLeuLeuAspAsnLysValGI	427	QY	708	lAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrVallysAlaAspAsnL	728
Db	1143	TAAAAAGATGCAACTGAGGAAGATAAAGAGAACCGTGAAATGAAAAATGCCAA	1202	Db	2036	TTAAC	2044
QY	427	nHisTyrSerPhePheGlyValGluGlnAlaAsnTrp	439	QY	728	eULeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr	747
Db	1203	AGCCTACCGTGTGACCAACCGACCAACACCATACCGCGGTATATCGAAGCCATTCA	1262	Db	2045	TGTTTCGACAGTCTACTATCCGACGACCAAGCGTGACCAATACCTCGCGGCGTG	2104
QY	440	-----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerI1	457	QY	747	lyArgSerPheThrGlyGlyValAsnVallysPhe	758
Db	1263	CGAGATTACGGCTTTACCTGACCGCGGCTGGCTTACGACCGCTTCAAGGTGAAAC	1322	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	457	eArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys	476	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1323	CCAC	1363	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	476	pLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheT	496	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1364	GTTC	1405	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	496	hrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnG	516	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1406	GCAGGAGCCACAACTAC	1450	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	516	luLeuTyrAlaHisGlyLysHis	532	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1451	CGCTGCAACCCAGCGCAACCGCGGATCATCTCGATTCCGACGGCAC	1500	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	532	lyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluG	552	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1501	-----AAAGCCGACCGCGCGCAATACCAATCGCTTCACTACAAAG	1546	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	552	lyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg	563	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1547	ACGGCAGCTTTCCGCAACCGCGATCTCTCCGCGACACCAACAGCGCTTGCCA	1606	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	564	-----AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly	578	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1607	ATCCGCAAAACCGCACGACTGTGCGCGTCCGGAAGCGGTCAACGCGCTACATCA	1666	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	578	-----	578	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1667	AAACCCAGGTWACGAATTGGCGGTCTCTCCGCGACCGCGCGCTGACCGCAAGTCG	1726	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	579	-----ArgGlyProLysSerIleGluAspSerGluMetLysLeuVala-gTyrA	596	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1727	CGGTAAGCGCGCAACACCGCGCTTTACGATACCCATCCTTAAACAACTGTTGAGCGCA	1786	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	596	snGlnSerGlyAlaAspPheTyrGlyAlaGlu--GlyGluIleTyr	610	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1787	ACCCGAG	1834	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	611	-----PheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG	628	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1835	CCTACCGCTTCAAA	1887	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	628	lyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspLysAlaTyrGlyAsnArgProp	648	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1898	-----CAAAAGCTACGGGTTCGATATTGG	1912	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	648	heIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisL	668	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1913	CGGAGGGCAAAAGACCGCGCAATTTGGAACCTGTAGCCAGGTTTCGGTG	1972	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	668	eUlysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaG	688	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1973	TGAAC	1990	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
QY	688	lnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyA	708	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139
Db	1991	ACTGGAACCGCTGGGCAAGACACGCTCAAT	2035	Db	2105	GACGTGATGTACGCTCGCGTGGCGTGAACCTACAAGTTC	2139

Alignment Scores:

Pred. No.:	6,05e-10	Length:	2368
Score:	223.00	Matches:	193
Percent Similarity:	33.45%	Conservative:	99

Key

Key	Location/Qualifiers
CDS	160...2302
/tag=	a
/product=	frpB
/note=	"70 kD major iron-regulated outer-membrane protein"

WO9631618-Al.

10-OCT-1996.

08-APR-1996; 96WO-US04774.

07-APR-1995; 95US-0418964.

(UYNC-) UNIV NORTH CAROLINA.

Beucher M, Sparling PF;

WPI; 1996-465038/46.

P-PSDB; AAW05181.

DNA encoding Neisseria iron-regulated outer membrane protein FrpB - useful for producing vaccine to protect mammals against N. gonorrhoeae and N. meningitidis infection

Claim 3; Fig 10; 68pp; English.

This sequence encodes a novel iron-regulated outer-membrane protein FrpB which was isolated from the gonococcal microorganism Neisseria gonorrhoeae strain FA1090. This gene is also homologous to one found in Neisseria meningitidis. The FrpB protein is surface exposed and immunogenic and is predicted to be useful as a vaccine because of its surface exposure, partial antigenic conservation and susceptibility to attack by bacterial antibodies. Such a vaccine will be of use against infection in mammals by N. gonorrhoeae and N. meningitidis. The protein and its fragments will also be useful in the diagnosis of Neisseria type infections.

Sequence 2368 BP; 685 A; 709 C; 542 G; 432 T; 0 other;

QY 596 snGlnSerGlyAlaaspPheTyrGlyAlaGlu---GlyGluLeuTyr----- 610
 Db 1945 ACCCGAG-----TTGGCGCACACACCGCGCGCACTTGGACGGCGCTCCCTTG 1992
 QY 611 -----PhelysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG 628
 Db 1993 CCTACCGCTTCAAA---AACCCGAATCTGMAATCGCTGGCGGACGCTATGTT--- 2045
 QY 628 lyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProp 648
 Db 2046 -----CAAAAAGCTACGGGTTCGATATTGG 2070
 QY 648 heileAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisL 668
 Db 2071 CGCGAGGCAAAAGACCGCGCAATTTGAAAACGTTGTACGCCAAGGTTTCGGTG 2130
 QY 668 euLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaG 688
 Db 2131 TGAAC-----GATGCTTCGCCA 2148
 QY 688 lnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisMetLeuAsnLeuGlyA 708
 Db 2149 ACTGGAACCGCTGGCAAGACACGCTCAAT-----GTTAATCTTTCGG 2193
 QY 708 laAsnTyrArgArgAsnThrArgTyrGlyGluTyrPAsnTrpTyrValLysAlaAspAsnL 728
 Db 2194 TTAAC-----AACG 2202
 QY 728 euLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetG 747
 Db 2203 TGTTCACAAAGTTCTACTACTCCGACAGCAACGCTGGACCAATACCTCGCGGGGTGG 2262
 QY 747 lyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2263 GACGTGATGTCGCTGGCGGTGAACACTACAAGTTC 2297

RESULT 34
 AAD10202
 ID AAD10202 standard; DNA; 2381 BP.
 XX
 AC AAD10202;
 XX
 DT 24-SEP-2001 (first entry)
 DE N. gonorrhoeae or N. meningitidis strain FA1090 FrpB protein DNA.
 XX
 DE Fe-regulated protein B; FrpB; outer-membrane protein; vaccine; infection;
 KW antibacterial; ds.
 XX
 OS Neisseria gonorrhoeae.
 OS Neisseria meningitidis.
 FH Key Location/Qualifiers
 CDS 172..2313
 FT /*tag= a
 FT /product= "N. gonorrhoeae or N. meningitidis
 FT strain FA1090 FrpB protein"
 XX
 PN US6265567-B1.
 XX
 PD 24-JUL-2001.
 XX
 PF 05-APR-1996; 96US-0628434.
 XX
 PR 07-APR-1995; 95US-0418964.
 XX
 PA (UYNC-) UNIV NORTH CAROLINA.
 PI Sparling PF, Beucher M;
 XX
 DR WPI; 2001-450739/48.
 DR P-PSDB; AAE05408.
 XX

PT An isolated nucleic acid encoding a Fe-regulated protein B (FrpB),
 PT useful for production of a vaccine against infection by Neisseria
 PT gonorrhoeae or N. meningitidis -
 XX
 XX Claim 2; Column 29-34; 20pp; English.
 XX
 CC The present DNA sequence encodes Fe-regulated protein B (FrpB) isolated
 CC from Neisseria gonorrhoeae and Neisseria meningitidis (strain FA1090). The
 CC present DNA sequence also represents the combined nucleotide sequence
 CC from PUNCH319 and PUNCH325. The FrpB is a 70 kD major iron-regulated,
 CC outer-membrane protein common to N. gonorrhoeae or N. meningitidis.
 CC The FrpB has particular use as a vaccine that protects a mammal from
 CC N. gonorrhoeae or N. meningitidis infection. The FrpB may also be used
 CC to detect the presence of antibodies specific for N. gonorrhoeae or
 CC N. meningitidis in a sample.
 XX
 SQ Sequence 2381 BP; 690 A; 712 C; 543 G; 436 T; 0 other;
 Alignment Scores:
 Pred. No.: 6.09e-10 Length: 2381
 Score: 223.00 Matches: 193
 Percent Similarity: 33.45% Conservative: 99
 Best Local Similarity: 22.11% Mismatches: 291
 Query Match: 5.53% Indels: 292
 DB: 22 Gaps: 45

US-09-936-377-2 (1-758) x AAD10202 (1-2381)

QY 11 LeuSerIleLeuLeuIleAsnThrProLeu---LeuAlaGlnAlaHisGluThrGluGln 29
 Db 193 CTCAGCTCTCTCGCTCACACTTGGCGGGCTTTCACCGCGGCAAGAAATAATGCC 252
 QY 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGly 49
 Db 253 AATGTCGATTGGATACCGTTACCGTAAAGGCGCGCCAA----- 294
 QY 50 LeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLys 69
 Db 295 -----GGCAGCAAAATCCGTACCAACATCGTT-----ACGCTTCAACAAAA 336
 QY 70 -----AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla 85
 Db 337 GACGAAGACCGCAACCGATATGCGCAACTCTTAAAGAGAGAGCCCTCCATC----- 390
 QY 86 SerGlnTyrGlyGlyGlyAlaSerAlaProValIle-----ArgGlyGln 100
 Db 391 ---GATTTCGGCGGCGCAACGCGACGCTCCCAATTCCTGACGCTGCGCGCATGGGTCAG 447
 QY 101 ThrGlyArgArgIleLysVal-----LeuAsnHisHis 111
 Db 448 AACTCTGTGCGCATCAAGGTGGACACACGCTATTCCGACAGCCAAATCCTTTACCACCAA 507
 QY 112 GlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAla 131
 Db 508 GGCAGA-----TTTATTGTCGATCCCGCT 531
 QY 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
 Db 532 TTGGTTAAAGTCGTTTCGTCACAAAAGGC-----GCGGTTCC 570
 QY 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu----- 169
 Db 571 GCCTCTGCGGTATCGCGCGCAACACGCGCGATTATCGCCAAAACCGTCGATGCCCAA 630
 QY 170 -----AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186
 Db 631 GACCTGCTCAAGGCTTGGATPAAACTGGGCGGTGGCCTCAACAGCGC----- 681
 QY 187 LysLeuThrSerGlyGlyIleAsnIleGly-----LeuGlyLys-----AsnPhe 201
 Db 682 TTTGCGGCAACACACGCGTAAAGTACGCGCGCAAGGTATTTCGAAAAGAGGGCACTTC 741
 QY 202 ValLeuHisThrGluGlyLeuTyr-----ArgLysSerGlyAspTyrAlaVal 217

```
Db 742 -----GACGGTTGTTCTCTTACACCGCAACGATGAAAGATTACGAAGCC 789
Qy 218 ProArg---TyrArgAsnLeu-----LysArgLeuProAspSerPro----- 230
Db 790 GGCAAGGCTTCGCAATGTCAACGGCGGCAAAACCGTACCGTACAGCGCGCTGACAA 849
Qy 231 -----ArgArgPheAlaAsnGlyGlnHisArgAlaValLeu 242
Db 850 CGCAGCTACCTCGCAAAATCGGAACAACCTTCGCGACGCGGACCCACCGCATCGTATTG 909
Qy 243 GlyTrpArgLysArgPheTyrArg-----ArgThrTyrSerAspArgAspGlnTyr 260
Db 910 AGCATATGAAGACCAACACCGGGGATCGGACTGTG-----CGTGAAGAGTTT 960
Qy 261 GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLys 280
Db 961 GCCGTGCGCGCGGAAAT-----TCA 981
Qy 281 SerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVal 300
Db 982 CGGATACTATTAAACGCCAAGCCCTCGCTACCGCGAA---ACCACACATCCACACACC 1038
Qy 301 AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 320
Db 1039 AATTGGCGTACACCGGCAAGAGATTGGCTTTCTCGAATACTGGATGCAACGCCCTAT 1098
Qy 321 AlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGlu 340
Db 1099 GTGCT---GGAAAAAAGACGCTATTCCGCGATGACAAAGATACCGCTACGCGAGCAA 1154
Qy 341 TrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
Db 1155 TGTAAGGCGCCNACCA-----TACCCGAATCGCCACTCGGG 1193
Qy 361 HisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsnGlnThrGlnAsn 380
Db 1194 CAT----- 1196
Qy 381 AlaArgIleGluLeuArgHisGlnPro-IleGlyArgLeuLysGlySerTrpGlyValGI 400
Db 1197 ---GAACTTCACTTCACAGCGCGCTTGCAGAACAAACCGCTGTGAAATACGCGCATCA 1253
Qy 400 nTyrLeuGlyGlnLys-----SerSe 407
Db 1254 CTACCGCCATCAGGAATCAACCGCAAGCGTGTGTTGAATTCACATTTAAATTGAAGA 1313
Qy 407 rAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGI 427
Db 1314 TAAAAAAGATGCAACTGAGGAAGATAAAAGAGAACCGGTGAAAAATGAAAAATGCGCAA 1373
Qy 427 nHisTyrSerPhePheGlyValGluGlnAlaAsnTrp----- 439
Db 1374 AGCTACCGTCTGACCAACCGGACCAAAACCGATACCGGCGGTATATCGAAGCCATTCA 1433
Qy 440 -----AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerII 457
Db 1434 CGAGATTGACGGCTTTACCTGACCGCGCGCTGCGTTACGACGCTTCAAGGTGAAGAAC 1493
Qy 457 eArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys---GlnProLeuProAs 476
Db 1494 CCAC-----GACCG-CAAAACCGGTTTCAAGCAGCAGCGCTCAACCCGA 1534
Qy 476 pLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheT 496
Db 1535 GTTT-----CGGCGTGATTGGCAGCGCGCGCAACACTGGAGCTTCA 1576
Qy 496 hrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnG 516
Db 1577 GCGGAGGCCAACACTAC-----GCCAGCGCGAGCGCGCGCTG-----TATGACG 1621
Qy 516 LeuLeuTyrAlaHisGlyLysHis-----ValAlaThrAsnThrPheGluValG 532
```

```
Db 1622 CGCTGCAAAACCCACGGCAAAACGGCGCATCATCTCGATTGCGACGGCACC----- 1671
Qy 532 lyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluG 552
Db 1672 -----AAAGCCGACGCGCGGCATATACGAAATCGGCTTCAACTACACAG 1717
Qy 552 lyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg----- 563
Db 1718 ACGCACGTTGTCGCAAAACGGCAGCTACTTCCGGCAGACCATCAAAAGACGCGCTTGCCA 1777
Qy 564 -----AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
Db 1778 ATCCGAAAAACCGCCACGACTCTGTGCGGTCGCGAAGCGTCAACGCGGCTACATCA 1837
Qy 578 ----- 578
Db 1838 AAAACACCGTTACGAATTGGCGCGTCTTACCACCGCGCGCTGACGCCAAAGTCG 1897
Qy 579 -----ArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrA 596
Db 1898 GCGTAAGCCGCAAAACCGCGCTTTTACGATACCCATCTCTAAAAAACTGTTGAGCGCA 1957
Qy 596 snGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyGluIleTyr----- 610
Db 1958 ACCCGAG-----TTTGGCGCACAAACCGCGCCGCACTTGGACGCGCTCCCTTG 2005
Qy 611 -----PheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG 628
Db 2006 CCTACCGCTTCAAA---AACCCGAATCTGGAATCGCTGGCGGAGCGTATGTT--- 2058
Qy 628 lyArgLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProp 648
Db 2059 -----CAAAAAGCTACGCGTTCGATATTGG 2083
Qy 648 heIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisL 668
Db 2084 CGCAGCGGCAAAAGACCGCGACGCAAAATGGAAAACGTTGTACGCCAAGTTTCGGT 2143
Qy 668 eulysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaG 688
Db 2144 TGAAC-----GATGCTCTGCCCA 2161
Qy 688 lnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyA 708
Db 2162 ACTGGAACCGCTGGGCAAAAGACACGCTCAAT-----GTTAACTTTTCGG 2206
Qy 708 laAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnL 728
Db 2207 TTAAC-----AACG 2215
Qy 728 euleuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetG 747
Db 2216 TGTTGACAAGTTCTACTATCCGCACAGCCACGCTGGACCAATACCTCGCGGCGTGG 2275
Qy 747 lyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2276 GACGTGATGTACGCTGGCGGTGAACACTCAAGTTTC 2310
RESULT 35
ABS78887
ID ABS78887 standard; DNA; 32160 BP.
XX
AC ABS78887;
XX
DT 17-DEC-2002 (first entry)
XX
DE E. coli CF7073 genomic sequence #54.
XX
KW Pathogenic; Escherichia coli CF7073 infection; livestock; pyelonephritis;
KW urinary tract infection; open reading frame; ORF; uropathogenic;
KW antibacterial; atrophic; nephrotropic; gene; ds.
XX
OS Escherichia coli.
```

XX WO200259320-A2.
 XX 01-AUG-2002.
 XX 19-OCT-2001; 2001WO-US46833.
 XX 19-OCT-2000; 2000US-242412P.
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 XX Blattner FR, Welch RA, Burland VD;
 XX WPI; 2002-691532/74.
 XX New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
 XX useful for preventing or treating E. coli CFT073 infection in humans or
 XX livestock -
 XX Claim 1; Page 166-184; 765pp; English.
 XX The present invention relates to polynucleotide sequences from the
 XX genome of the pathogenic Escherichia coli strain CFT073. Almost all
 XX the sequences present in E. coli CFT073 are absent in the previously
 XX sequenced laboratory strain K-12. The polynucleotide sequences of
 XX the invention are useful for preventing, diagnosing or treating
 XX E. coli CFT073 infection in humans or livestock. The polynucleotide
 XX sequences are useful for preventing urinary tract infections and
 XX pyelonephritis. Likewise, the polypeptides encoded by the different
 XX open reading frames (ORF1-5) are useful for generating a vaccine
 XX against uropathogenic E. coli strains. ABS78834-ABS79085 represent
 XX genomic sequences from E. coli strain CFT073.
 XX SQ Sequence 32160 BP; 7001 A; 9686 C; 8766 G; 6688 T; 19 other;
 Alignment Scores:
 Pred. No.: 3,71e-08 Length: 32160
 Score: 220.50 Matches: 187
 Percent Similarity: 34.05% Conservative: 97
 Best Local Similarity: 22.42% Mismatches: 309
 Query Match: 5.46% Indels: 242
 DB: 24 Gaps: 45
 US-09-936-377-2 (1-758) x ABS78887 (1-32160)
 QY 1 MetAlaGlnThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuLeuLeuLeu 20
 DB 29138 ATGAAATACACACGGCTTTATCCTCTGGCTTGGGGGATATTGCTC-----CCCGCC 29191
 QY 21 LeuAlaGlnAlaHisGlnThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
 DB 29192 ATTGCTAATGCCAGACTTCACAGCAGAC-----GAAGCAGCGCTGGGTATCC 29242
 QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
 DB 29243 GCCAGTAAACAATCTCCCGCTCGGCA-----TCAGCCACAACAGCTCTCATCTGTT 29296
 QY 61 IleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
 DB 29297 GTCAGCGCCCGGAATATACGACGCGCGCTCACCGCAGGACAAACTCCCGCAGATC 29356
 QY 81 ---ProGlyIleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProVal---IleArg 98
 DB 29357 TTGCCCGGGCTCAATATTGAAATACGCGCAACATGCTTTTTCGAGATCTCGCTACGC 29416
 QY 99 GlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAla 118
 DB 29417 GCGGCTCTCTTCAGCGCAG----- 29434
 QY 119 AspPhe---SerProAspHisAlaIleMetValAsp----- 129
 DB 29435 GACTTCTATACCCCGCGCTCACCCCTGTATGTCGATGGCGTCCCTCAGCTTCCACCAAC 29494

QY 130 -----ThrAlaLeuSer-----GlnGlnValGluLeuLeuArgGlyProValThrLeu 145
 DB 29495 ACCATCCAGCGCGCTTACCGATGTGCAAGCGTGGAGTTGCTGCGAGGCCACACGGAACG 29554
 QY 146 LeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGlu 165
 DB 29555 TTATATGGCAAAAGCGCTCAGCGCGGATCATCAATCGTCCACCGCAG--CCGGAC 29611
 QY 166 LysMetPro-----GluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSer 181
 DB 29612 AGCAGCGCGCGCTATATTGAAGCGCGCTCAGTAGCCGACAGTATTCTGA---AGT 29668
 QY 182 SerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPhe 201
 DB 29669 AAGTTCAACCTG-----AGCGGCCCATTCAGGATGCCTG----- 29704
 QY 202 ValLeuHisThrGluGlyLeuTyrArgLys-----SerGlyAspTyrAlaValProArg 219
 DB 29705 CTGTACGCGACGCTCACCTGTTACGCCAGGTTGATGACGCGCAGCATGATTAA-CCCGCG 29763
 QY 220 -----TyrArgAsnLeuLysArgLeuProAspSerProArg-----ArgPheAlaAsn 235
 DB 29764 GACGGGAAGCGATGACTTAGCGCGCACCCGCGCAGCATAGGAATGAACTGCGTCT 29823
 QY 236 GlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAsp 255
 DB 29824 GCGCGCGACGATCAGCCCTGGGA----- 29847
 QY 256 ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAsp 275
 DB 29848 -----AATGGGCTT-----TGC----- 29859
 QY 276 IleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeu 295
 DB 29860 -----CGCTCAGCGAATGATACCCGCG----- 29883
 QY 296 ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys-GlyPheHisAspAspAs 315
 DB 29884 -----CAACCGAGGATGCTTGTGGGATGGATGATATTAA 29919
 QY 315 pAspAlaHisAlaHisAlaHisAsnGlyLysPro----- 326
 DB 29920 GGGCCGTAAGCTGTGATCAGCGATGTTTACCAGACCCGTATACGCGCGCTGCACGTA 29979
 QY 327 -----TyrIleAspLeuArgAsnLysAr 334
 DB 29980 CAGCCAGACCTCAGTGGGAAATACACACCGATGACTGGT----- 30022
 QY 334 gTyrGluLeuArgAlaGluTyrLysGln-----ProPheProGlyPheG1 349
 DB 30023 -TTCAACCTGTATCAGCGCTGGCAGCAGCAGCATATTTCGCGCACCTTCCCTTCC---GG 30078
 QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLys----- 365
 DB 30079 TTCGTTAATCGTCAATATCCCTCAGCGC---TGGAAATCAGGATGTGACGAGCTGCGCGC 30135
 QY 366 -----AlaGlyAspAla-----ValGluAsnPhePheAsnAsnGlnThrGlnAsnAl 381
 DB 30136 CGCAACCTCGGCGGATGCGCTGATATGTTGTTGGCTGTACCGCAGAACAC 30195
 QY 381 aArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyr 401
 DB 30196 CCGC-----GAGAAGTTAAATTCAGCTACGACATGCGGAC 30231
 QY 401 rLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLe 421
 DB 30232 AATGCTTATTAAACAGTACCGGCTATACCCAGCTGAAACGCTGGCGCA----- 30283
 QY 421 uLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp----- 439
 DB 30284 -----TACAGT-----GACCTGACCTGGCATTT 30306
 QY 440 -----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleAr 458

Db 2113 AGTGTTACCTACCAAGTTATATACTGACCGAATCTCGTGCTACCTATGCG-----CCA 2116
 QY 715 ArgTyrGlyGluTrp 719
 Db 2167 TTAAGAGCGGAATGG 2181
 RESULT 37
 AAZ54329
 ID AAZ54329 standard; DNA; 2112 BP.
 XX
 XX AAZ54329;
 XX
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 761 partial DNA sequence SEQ ID NO:2607.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 01-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PA
 PI Patersen C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR P-PSDB; AAY75567.
 XX
 DR Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 7; Page 1236-1237; 1453pp; English.
 XX
 CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitis and N. gonorrhoea polynucleotides
 CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 2112 BP; 555 A; 650 C; 484 G; 423 T; 0 other;

```

QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAla 319
Db : : : : :
880 CTGGAATAC----- 888
QY 320 HisAlaHisAsnGlyLysProTrrPileAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db : : : : :
889 ---GCCTTCAACGACAAA---TGG-----CGTGCC 912
QY 340 GluTrrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db : : : : :
913 CAATGGCAG-----CTGCCCCACCGCACGGCGGCAGGAT 948
QY 360 TyrHisHis-----AspGluLysAlaGlyAspAlaValGluAsnPhenPhe----- 374
Db : : : : :
949 TTTGATCATTTCTATGTCAGGCAGCAGCAAAATGGCAACTTATCAAAAGTAACTACGCGCTCG 1008
QY 375 -----AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArg---HisGlnProIle 390
Db : : : : :
1009 CAGCAGACCGACACAAACACCTGCTGCTCCAACTTAACGCTCAACGGCGACTACACCAATC 1068
QY 391 GlyArgLeuLysGly-----SerTrrGlyValGlnTyrLeuGlyGlnLys----- 405
Db : : : : :
1069 GCGCGTTTGGAAACACCACTGACCGTAGGCATGGATTACAGCGCGCAACACCGCAACCCCG 1128
QY 406 -----SerSerAlaLeuSerAlaThr----- 412
Db : : : : :
1129 ACATGGGTTTCACAGCGGCTTTTCGCTCTCATCAACCCCTACGACCGCGCAAGCTGG 1188
QY 413 ---SerGluAlaValLysGlnProMetLeuLeuAspAsn-----LysValGlnHisTyr 429
Db : : : : :
1189 CCGGCTTCGGCAGATTGCGACCTTATCTGACCCCAAAACCGCACAAAGCGCACTCCTAC 1248
QY 430 SerPhePhe-----GlyValGluAlaAsnTrpAsp---AsnPhenThrLeuGluGly 446
Db : : : : :
1249 GGCATCTTTGGCAAAACATCTCTCCGCGACGCGCGGATTTGAAATTCGCTCGCGGCG 1308
QY 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLysAspArg 466
Db : : : : :
1309 -----CGTTTACGACAAATACACCTTTAATTC 1335
QY 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln---ThrAlaArg 485
Db : : : : :
1336 GAAACAAACATC-----ACCGGCGACGACGCGCCCAATACAGCGGACAC 1377
QY 486 SerPheAla-----LeuSerGlyAsnTrrPyrPheThrProGlnHisLysLeuSerLeu 503
Db : : : : :
1378 TCGTTACCCCAACATCGCGGAGTGTGGACATCAATCCCGTCCAC-----ACACTT 1431
QY 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
Db : : : : :
1432 TACGCTCGTATAACAAA-----GGCTTCGCGCTTATGGCGGACGCGCGGCTAT 1482
QY 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db : : : : :
1483 TTGAGCATCATACGTGCTCTCCGCGGTGTTCAACCGCCGACCCCGAGTACCCCGCCAA 1542
QY 544 IleGluLeuAlaLeu-----GlyTrrGluGlyAspArgTrrPrrGlnTyrAsnLeuAlaLeu 561
Db : : : : :
1543 TACGAAACCGGCGTGAAGAGCAGTTGGCTGGACGACCGCTCAGCACTACGTTGCTGCTGCC 1602
QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
Db : : : : :
1603 TAC----- 1605
QY 582 LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
Db : : : : :
1606 -----CAAATCGACGCTTCAT----- 1623
QY 602 PheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArg-----TyrArg 618
Db : : : : :
1624 -----ATCGCTACCGCGCGGATCCAAAAAACAACCCCTATATT 1662
QY 619 IleGlyValSerGlyAsp----- 624

```

```

Db 1663 TATGCGGTTAGCGGCAAAACACCGTTCCGCGCGGTGGAATTTGTCGCCCATCGGCAATC 1722
QY : : : : :
625 -----TyrValArgGlyArgLeuLysAsnLeuLysAsnLeuProSerLeuProGlyArg 639
Db : : : : :
1723 ATCCCAAAAAACTCTATCTCGCGGTTCGTTGGCGGTGATGACGCGGAAA----- 1773
QY 640 GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
Db : : : : :
1774 -----GTCGTTGAAGACAAGAAATCCCGACCGAGTG 1806
QY 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
Db : : : : :
1807 -----GGCATCCATTG-----ATAATACCAACGATTACCGGCAAC 1845
QY 680 LeuAspTyrTyrArgVal----- 685
Db : : : : :
1846 CTG---TTTTCGTTATATACCCCGACGAAACCTCTACGCGGAAATCGGCGTAACCGGT 1902
QY 686 -----PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHis 701
Db : : : : :
1903 ACAGCAAAACGCTACCGTTACAACTCAAGAAATAAAGAGTGACTACGCTTCCAGGCTTT 1962
QY 702 HisMetLeuAsn-----LeuGlyAlaAsnTyrArgArg---AsnThrArgTyrGlyGlu 718
Db : : : : :
1963 GCGCGAGTTGATGCCATCTTGGCTGGAAACCAATAAAATGTTAACGTTACCTTGC--- 2019
QY 719 TrpAsnTrrPyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSer 738
Db : : : : :
2020 -----CGAGCAATCTGCTCAATCAAAATATTTGGCGTTTCGGACTCT 2061
QY 739 PheLeuSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db : : : : :
2062 ATG-----CCGGTAAATCCGCGGCTATATCTGCCCGGTAAATTACCGTTTC 2109

RESULT 38
AAF56467
ID AAF56467 standard; DNA; 2112 BP.
XX
AC AAF56467;
DT 18-APR-2001 (first entry)
XX
DE Neisseria meningitidis coding sequence #26.
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
XX rni5; rth; tolC; ds.
XX
OS Neisseria meningitidis.
XX
PN EP1069133-A1.
XX
PD 17-JAN-2001.
XX
PF 13-JUL-1999; 99EP-0401764.
XX
PR 13-JUL-1999; 99EP-0401764.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Nassif X, Tinsley C;
XX
DR WPI; 2001-082916/10.
DR P-PSDB; AAB68927.
XX
PT Immunogenic polypeptides derived from Neisseria meningitidis and the
PT nucleic acids that encode them, useful for diagnosing and vaccinating
XX against Neisseria infections e.g. bacteremia and meningitis -
XX
PS Claim 11; Fig 26A; 240pp; English.
XX
CC The present invention provides the protein and coding sequences of

```



```
Db 7269 GCTACGATATCGCGGTGAAGC-----ATTTCTCGCGGTTTT 7228
QY 114 ThrGlyAspMetAlaAspPheSerProAspHisAla-----IleMetVal 128
Db 7227 CAAGCGGACGCATCCGATATTATTCGCGAGCGCGTGGGAAAGCGGACAAGTGGCGCG 7168
QY 129 AspThrAlaLeuSerGlnGlnValGluLeuLeuArgGlyProValThrLeuLeuTyrSer 148
Db 7167 AGTACTGCCAACATCAGGCGGTGGAATCCCTGAAGGCCCGCTTTCCGTGCTTTACGGC 7108
QY 149 SerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetPro 168
Db 7107 CGCAACACGCGCGCGGTCTACATCGTTCAGCAATATCGCCAACTTCAACAAAGC 7048
QY 169 GluAsn-----GlyValSerGly----- 174
Db 7047 CGCAACATCGGAGCGGTTTACGGCTCATGGGCAACCGCAGCGCTGAATATGACATTAAC 6988
QY 175 -----GluLeuGlyLeuArgLeuSerSer-----GlyAsnLeuGluLys 187
Db 6987 GAAGTGTGCAACAAACATCCGCCATCCCTCACCAGCGAAGTCGGCGGCCCAATTGC 6928
QY 188 LeuThrSerGly-----GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThr 205
Db 6927 TTCCGCGCGGATAGACAGCAAAATGTCATGGTTTCGCCAGCATTAACCTCAAACTC 6868
QY 206 GluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225
Db 6867 GACAACGGCTGAAGTGGAGGGGCAATACACC-----TAGCACAATGTGGAGCGC 6817
QY 226 LeuProAsp---SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrp 244
Db 6816 ACGCCGACGCGAGTCGG----- 6799
QY 245 ArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuPro--- 263
Db 6798 ACCAAGTCCGTGTAC-----GACCCTTCGGACTCGCTTAC 6763
QY 264 -----AlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
Db 6762 CGCATGGGGTTCGCCCAACCGGAACGATTT----- 6733
QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLysAsp 299
Db 6732 -----GTCAAGACAAAGTGCAGTTTGG-----CGTTCCGAC 6700
QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
Db 6699 CTTGAATAC----- 6691
QY 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db 6690 ---GCCTTCAACGACAAA---TGG-----CGTCC 6667
QY 340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db 6666 CAATGGCAG-----CTCGCCCAACCGCAGCGCGCGGAGAT 6631
QY 360 TyrHisHis-----AspGluLysAlaGlyAspAlaValGluAsnPhePhe----- 374
Db 6630 TTTGATCATTCTTATGCGAGCGCAGCAAAATGGCACTTAATCAACGTAACCTACGCCCTGG 6571
QY 375 -----AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArg---HisGlnProIle 390
Db 6570 CAGCAGCCGACACAAACCTCTGCTGCCAACTTAACGCTCAACGGCGGACTACACCATC 6511
QY 391 GlyArgLeuLysGly-----SerTrpGlyValGlnTyrLeuGlyGlnLys----- 405
Db 6510 GGCGGTTTTGAATAACACCTGACCGTACCGTACGATGATTAACAGCGCGAACCAGCCG 6451
QY 406 -----SerSerAlaLeuSerAlaThr----- 412
```



```
Db 302426 CGCATGGGTTCCGCCACCGAAGATT--- 302455
QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
Db 302456 -----GTCAAAGACAAGCTGCAAGTTTG-----CGTTCGGAC 302488
QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
Db 302489 CTTGATATAC----- 302497
QY 320 HisAlaHisAsnGlyLysProTyrPheAspLeuArgAsnLysArgTyrGluLeuArgAla 339
Db 302498 ---GCCTTCACGACAAA---TGG-----CGTGGC 302521
QY 340 GluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
Db 302522 CAATGGCAG-----CTCGCCACCGCACGCGCGGCGCAGAT 302557
QY 360 TyrHisHis-----AspGluLysAlaGlyAspAlaValGluAsnPhePhe----- 374
Db 302558 TTTGATCATTTCTATGCAGCGAGGAAATGGCACTTAATCAACGTAACTACGCCCTGG 302617
QY 375 -----AsnAsnGlnThrGlnAsnAlaArgIleGluLeuArg---HisGlnProIle 390
Db 302618 CAGCAGACCGACACAAACCTCTGCTCCAACCTTAACGCTCAACGGCGACTACACCATC 302677
QY 391 GlyArgLeuLysGly-----SerTyrGlyValGlnTyrLeuGlyGlnLys----- 405
Db 302678 GGCCTGTTTGAACACCAACCTAGCAGTGGATGATTACAGCCGCGAACAACCCGCAACCCG 302737
QY 406 -----SerSerAlaLeuSerAlaThr----- 412
Db 302738 ACATTGGGTTTCAGCAGCGCTTTTCGCTCCATCAACCCCTACGACCGCGCAAGCTGG 302797
QY 413 ---SerGluAlaValLysGlnProMetLeuLeuAspAsn-----LysValGlnHisTyr 429
Db 302798 CCGGCTTCGGCGCAGATTGCGAGCCTATTCTGACCCAAACCGCCACAAAGCCGACTCTCTAC 302857
QY 430 SerPhePhe-----GlyValGluGlnAlaAsnTyrAsp---AsnPheThrLeuGluGly 446
Db 302858 GGCATCTTTGTGMAAACATCTCTCCGCCACGCCCGGATTTGAAATTCGTCGCGCGGC 302917
QY 447 GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArg 456
Db 302918 -----CGTTACGCAAAATACACCTTTAATTC 302944
QY 467 GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln---ThrAlaArg 485
Db 302945 GAAACACAAACTC-----ACCGCAGCAGCGCGCCCAATACAGCGGACAC 302986
QY 486 SerPheAla-----LeuSerGlyAsnTyrPheThrProGlnHisLysLeuSerLeu 503
Db 302987 TCGTTTCAGCCCCCAACATCGCGCAGTGTGAACATCAATCCGTCCTAC-----ACACTT 303040
QY 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
Db 303041 TACGCTTCGTATAACAAA-----GGCTTCGCGCTTATGCGCGACGCGCGGCTAT 303091
QY 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db 303092 TTGAGCATCGATAGTTGTTCTTCGCGGTTCACAGCCGACCCCGAGTAGTACACCCGCCAA 303151
QY 544 IleGluLeuAlaLeu-----GlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeu 561
Db 303152 TACGAAACCGCGGTGAAAGCAGTGTGGCTGGACCGCTCAGCACTAGCTTGTCTGGCC 303211
QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
Db 303212 TAC----- 303214
QY 582 LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
Db 303215 -----CAAATCGAACGCTTCAAT----- 303232
```

Search completed: November 17, 2003, 18:17:17

Job time : 3725 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 17:11:51 ; Search time 99 Seconds
(without alignments)

3379.479 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTTLKPIVLSILLINTEPL.....FLSDTFQMRSGTGGVNVKF 758

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09936377/runat_14112003_104450_22141/app.query.fasta_1.903
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09936377@cgn_1_1_44@runat_14112003_104450_22141 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFPLOCKS=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match %	Length DB ID	Description	
1	1391	34.5	2115	4	US-09-328-352-1018
2	685	17.0	2157	4	US-09-252-991A-12861
3	405.5	10.0	1395	4	US-09-252-991A-13078
4	319.5	7.9	2226	4	US-09-328-352-1772
5	255	6.3	2262	4	US-09-252-991A-1416
6	239.5	5.9	2370	4	US-09-252-991A-10440
7	230	5.7	2169	5	PCT-US96-05320A-264
8	230	5.7	1830121	4	US-09-557-884-1
9	230	5.7	1830121	4	US-09-643-990A-1
10	224.5	5.6	1827	4	US-09-252-991A-1375
11	223	5.5	2381	3	US-08-628-434-3
12	219	5.4	2232	5	PCT-US96-05320A-113

ALIGNMENTS

RESULT 1

US-09-328-352-1018
; Sequence 1018, Application US/09328352
; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Berton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 1018

; LENGTH: 2115

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-09-328-352-1018

Alignment Scores:
Pred. No.: 5.7e-135 Length: 2115
Score: 1391.00 Matches: 315
Percent Similarity: 56.19% Conservative: 121
Best Local Similarity: 40.59% Mismatches: 222
Query Match: 34.46% Indels: 118
DB: 4 Gaps: 24

US-09-936-377-2 (1-758) x US-09-328-352-1018 (1-2115)

QY	9	lleValLeuSerIleu---LeuIleasnThrProLeuLeuAlaGluAlaHisGluThr	27
DB	61	TTGACCTTATCATTTTGGCTGCCATATCTATCTGTCAGCGAAATGAAAT	120
QY	28	GluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThr	47
DB	121	GTGAGAAG-----TTAGAACTATCCGATC-----AGGCCACCCC-----	159

Sequence 12930, A
Sequence 6426, Ap
Sequence 1, Appli
Sequence 1, Appli
Sequence 14350, A
Sequence 14267, A
Sequence 970, App
Sequence 1192, Ap
Sequence 708, App
Sequence 6, Appli
Sequence 194, App
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 14238, A
Sequence 5404, Ap
Sequence 2, Appli
Sequence 1, Appli
Sequence 1816, Ap
Sequence 174, App
Sequence 1708, Ap
Sequence 2016, Ap
Sequence 9244, Ap
Sequence 15877, A
Sequence 14364, A
Sequence 10725, A
Sequence 8394, Ap

US-09-252-991A-12930
US-09-252-991A-6426
US-08-628-434-1
US-08-425-843-1
US-09-252-991A-14350
US-09-252-991A-14267
US-09-221-017B-970
US-09-252-991A-1192
US-09-328-352-708
US-08-425-843-6
US-09-453-702B-194
US-08-487-890A-4
US-08-478-435-4
US-08-337-483-4
US-08-478-373-4
US-08-474-671-4
US-08-483-577A-4
US-08-897-438-4
US-08-637-654-4
US-08-649-518-4
US-09-252-991A-14238
US-09-252-991A-5404
US-09-307-973A-1
US-09-328-352-1816
US-09-453-702B-174
US-09-252-991A-1708
US-09-252-991A-2016
US-09-252-991A-9244
US-09-252-991A-15877
US-09-252-991A-14364
US-09-252-991A-10725
US-09-252-991A-8394

Qy 48 SerGlyLeuLeuHisThrSer-----ThrAlaSerAspLysIleSerGlyAsp 64
Db 160 -----CTTGAGCAAACTTCAAAAGATTTTCAGTCGCTGATCTAGTTGACCAAAAG 213
Qy 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHis 84
Db 214 CATTTGACTGAAGGGCCGCCACTATTGGGATCGCTCAATAGTAGAGTGGGGATTTAT 273
Qy 85 AlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArg 104
Db 274 GCGAATCAGTTTGGGGCTGGTTCAGTCTCGGTGATTTCGTGGCCAAAGATGGACCACT 333
Qy 105 IleLysValLeuAsnHisHisGlyGlnThrGlyAspMetAlaAspPheSerProAspHis 124
Db 334 GTTAAGTCTTCAAAATTCATCTGAAATGTGGATGTCTCAACTTTGTCTCTCGATCAT 393
Qy 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThr 144
Db 394 GCTGTAAACAGTAGATCCAGTTCTTGCAAAACAAGTTGAAGTGAATTCGTGGACCTCTACT 453
Qy 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyIlePro 164
Db 454 TTATTTGTTGGTCAGGAACTGTGGGTGTTTGTCAACGTTATTTGATAACAAGATCCCT 513
Qy 165 GluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn 184
Db 514 ACACAAATGCCCTGAAATGGTTATGAAGCAAGTTGGCTTACGTTTCAACACATGGAAGT 573
Qy 185 LeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHis 204
Db 574 GATGAAAGCTGGCAAGCGCTGGCGTAACCTGTGGCTTGGCCAGCAAGTTGCTTTACGT 633
Qy 205 ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyr----- 220
Db 634 GTTGAGGGGTTAACTCGGTGATGCGCAATAATATATATGCTCCCAATATATATCATCAAGGT 693
Qy 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
Db 694 GAAAGAGCGCCGTGTAGACAATACT-----TTTGGCGAGGTGATCGGTAAAT 744
Qy 241 Val---LeuGlyTyrArgLysArgPheTyrArgArg-----ThrTyrSer 254
Db 745 GTTGGTTTGTCTTGG-----ATTATGACCGTGGTTATACGGGTATTTCTTATAGC 795
Qy 255 AspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlyTyrAspAspCysHisAla 274
Db 796 AACCGCAGACCACTAGTATGGCTTACCGCGCATAGCCATGATGACGAAAGTTGTAGCGCT 855
Qy 275 AspileIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeu 294
Db 856 -----CATCTA 861
Qy 295 LeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp 314
Db 862 -----GGTGAAGACCACTTATACATTTGATGATGCTCATGAGCAT 900
Qy 315 Asp-----AspAlaHisAlaHisAla-----HisAsnGlyLys 325
Db 901 GATCATGAAGGGTGAAGAGCTCATGCCCATGAGAACTATGACGACGACATGCGCGC 960
Qy 326 ProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPhe 345
Db 961 CCATGGATTGACTTAAATCTGACGGTTATGATTTTAAACCGATGTTAAATGATCCATTT 1020
Qy 346 ProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLys 365
Db 1021 GCAGGCTTCCAAAAGTTTACGTGCCCAAGCGCATATACTGACTATCAGCATGATGAAT 1080
Qy 366 AlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeu 385
Db 1081 GAAGAAGGCAATCGCGACTCGATTCCAAACAAAGGTTATGATGGCGGTATTGAGTTA 1140
Qy 386 ArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLys 405

Db 1141 GTCCATAATCTCTATTGCTGATTGGGAAGGGGTAAATGGTACCAG-----TAGGCCAACAA 1197
Qy 406 SerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLys 425
Db 1198 AAA-----CTAAATTTAACTGGTGAA-----GAAGCATTTTATGCCACCAACCACT 1242
Qy 426 ValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTyrAspAsnPheThrLeuGlu 445
Db 1243 ACCAAAAGTGGAGTGTTCCTTTAGAAATAGAAATGAAAGATGTTTCATTTGAA 1302
Qy 446 GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAsp 465
Db 1303 CTTTCTGCTGTCAGATCAGCAAGAA-----ATAGAT 1335
Qy 466 ArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArg 485
Db 1336 GTAGATGATAACTCTAAACAAGATTTTGACCGT-----TCTGCTTTT 1377
Qy 486 SerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSerLeuThrAla 505
Db 1378 TCCTATGTCAGGTGCTGCCAAGTGGGAGTTTGGCCAAATATATAGCTTTTCATTTGTAGCA 1437
Qy 506 SerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAla 525
Db 1438 TCACACCAAGAGCGTCTACCTCTTGCACAGAGTTATATGCAAAATGGTGCCTTTTTCGG 1497
Qy 526 ThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGlu 545
Db 1498 ACCAATCTTATAGTTAGGAAATGATCACTCAGTAAAGAAATCAAAATAATGTTGAG 1557
Qy 546 LeuAlaLeuGlyTyrGluGlyAspArgTyrPheThrPheThrPheThrPheThrPheThrPhe 565
Db 1558 TTAGGTTTGCATTTTGTATGATGACAACTCGAATTCATCTTCATGTTTACCAATACCTGG 1617
Qy 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGlu 585
Db 1618 TTTGATGACTACATCTATGCGCAAAACCTTGTACCGCTTATATA----- 1659
Qy 586 AspAspSerGlnMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla 605
Db 1660 -----GACTTCGGTTTAGTTCAATACACTCAAGATAAGTCGTTTCTATGGTGCT 1710
Qy 606 GluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyr 625
Db 1711 GAGGGTGAATTTGGTTATCAAAATACGCCAATGTACAAGATCAGTCTTTTGGCGATTAC 1770
Qy 626 ValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsn 645
Db 1771 GTGCGCGGTGAAGATT-----GATCGCGAGGCG----- 1797
Qy 646 ArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGly 665
Db 1798 -----AATGCTCTGTTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1830
Qy 666 PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
Db 1831 ACTAAAGTTGATGACAGATTTTGGTACCGTTTATAGTGGTTCTGCTGCTGCTGCTGCTGCT 1890
Qy 686 PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsn 705
Db 1891 TTTAATCAAGATAGATCGCAGCTTATGAAACCGAAACAGAGAGGCTACATCATCATCAT 1950
Qy 706 LeuGlyAlaAsnTyrArgArgAsnThrArgTyrGly-----GluTyrAsnTyrTyr 722
Db 1951 TTAGGGGTAGCATAC-----TCTGGCAATATGGTCCAAAACAGATATATCGGTTTAT 2004
Qy 723 ValIleAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAsp 742
Db 2005 ATGAAAGCAAAATAACTTGTATGATGACACCGTTTATCAACATCGCTGCTTCTTATCAAT 2064
Qy 743 ThrProGlnMetGlyArgSerPheThrGlyValValAsnValLysPhe 758

Db 375 ATCCGGACCTCGAGCGCGTTTCACTCGTGAAGCTGAGCCGCCCTATACCAAGTAC 316
 Qy 361 HisHsAspGluLysAlaGlyAspAlaValGluAsnPheAsnGlnThrGlnAsn 380
 Db 315 GAGCACAGGAATCGAGGATGCGGAGACCGGACCACTTCAAGAACGAGGCTACGAA 256
 Qy 381 AlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGln 400
 Db 255 GGCGCATCGAGCGCCGCCACCGCGCTCGCGCGCTGAACGGGCTGGTGGCGCGCAG 196
 Qy 401 TyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMet 420
 Db 195 TTCGCCAACGCGCTTCTCGGCCCTCGG-----GAGGAGCGCTTC 154
 Qy 421 LeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp--- 439
 Db 153 GTGCCGCACAGGAACCGACAGCGCGCGCTGTTCGCCCTGGAGAA-----TGGAAAG 100
 Qy 440 -----AspAsnPheThrLeuGluGlyGlyValArgValGlu 451
 Db 99 CTCAGCAGCGCTCGACCTCAGCTTCGCGCGCGCGCTGGAG 58
 RESULT 4
 ; US-09-328-352-1772
 ; Sequence 1772, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 1772
 ; LENGTH: 2226
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-1772
 Alignment Scores:
 Pred. No.: 3,68e-23 Length: 2226
 Score: 319.50 Matches: 164
 Percent Similarity: 36.38% Conservative: 127
 Best Local Similarity: 20.50% Mismatches: 333
 Query Match: 7.92% Indels: 176
 DB: 4 Gaps: 30
 US-09-936-377-2 (1-758) x US-09-328-352-1772 (1-2226)
 Qy 59 LysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAsp 78
 Db 46 AAGTAGAAGTCCCAATACA-----ATAGGTGATGCTCTTAAA 84
 Qy 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArg 98
 Db 85 AATATAACAGGTATACAAAGTACTTTCATTTGGACCCCAATGCAGGTGCACCGATTATACGT 144
 Qy 99 GlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAla 118
 Db 145 AGTTTGAGTGAATCGGTTGGAGTGAATGAAATCGAATTAATTAATGGAATGAAT 204
 Qy 119 AspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIle 138
 Db 205 GCATTTAGTGACATATTAATATACCATTCGATCCCAATTTTATAGAAAAGTATAGTAGT 264
 Qy 139 LeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspVal 158
 Db 265 AACAAAAATACAGATAATATTCGTTATGGCGGAAATGCAATAGGCGGAGTGTTCAAATA 324
 Qy 159 AlaaspGlyLysIleProGluLysMetProGluAsn----- 170
 Db 325 GAATCGGTTTAAATTCCTTAAATAATTGAAGAAAGCCAAACAATTAGATATTGCTCTTT 384

Qy 171 -----GlyValSerGly----- 174
 Db 385 CGGAAAGGATTTAATGATTTTGAATGTAAGGGGTCAATTTTAAATCAATGATCAAAAA 444
 Qy 175 -----GluLeuGlyLeuArgLeuSerSer-----Gly 183
 Db 445 AACTGGTCTCAAAATATAAGATATTTCAGAAATATGAATTTCTTTATAAAATCCAGGA 504
 Qy 184 AsnLeuGlu-----LysLeu-----ThrSerGlyGlyIleAsnIle 195
 Db 505 AATAGTAAGCCAATTTATGTGAAGCAGAGATTTTTCAAATTCAGGTGGGATCAATAGT 564
 Qy 196 GlyLeuGlyLysAsnPhe-----ValLeuHisThrGluGlyLeuTyrArg 210
 Db 565 GCATTAGCAGCTTCTGTCAAAAGATAGTAGTCCCAACAT-----ATTATATAC 615
 Qy 211 LysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerPro 230
 Db 616 AAGTCGTCAACACCATACATAGATAAATTT-----ATGACTGAGAATCCTGAT----- 663
 Qy 231 ArgArgPheAlaAsnGlyGln-----HisArgAlaValLeuGlyTrp--A 245
 Db 664 -----TGGGCAGATGGCGATTTTCTTTTATACAGATAAGCCACATCTATATGGGA 717
 Qy 245 IrgLysArgPheTyrArgArgThrTyrSer-----A 255
 Db 718 GGAAAAACATATATAAATCCAAAAATCCAGAGTACATACCTAATCTCCGCAAAATACA 777
 Qy 255 sPArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaA 275
 Db 778 ATAAAAAGATCAATACGGA-----TGTACACCA 807
 Qy 275 spile-----lleTrpGlnL 280
 Db 808 AATTATTTTAAAAAATTAGGTAAACAGTTATGCTCAGAAATGAGAATATTGGGTTCGAA 867
 Qy 280 ysSerLeu--lleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
 Db 868 ACATACCTTTTGTATAAGGTTTCATGGATTG-----AGTGCAGATAGAAAAA 918
 Qy 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
 Db 919 AGTGAATATGGTGTACCAAGATTTTCATTA-----CAAAATCAATCTTTTGCT 966
 Qy 320 HisAlaHisAsnGlyLysPro---TrpIleAspLeuArgAsnLysArgTyrGluLeuArg 338
 Db 967 GATTCTTATGAACAAATACCGGTGGTGTGAATAATAGATCAAAATCGTTTGTATTAAAT 1026
 Qy 339 AlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn 358
 Db 1027 TCCAAATTTTATCCAACTATTCCTCTGCAGAAAGAGATAAGTTTGAATTTTCAACAGCTT 1086
 Qy 359 AspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnGlnThr 378
 Db 1087 TCTAATAAATATCGGAGATATGTTGTTCTGTCGAAAGAGCTAATAGTATATAATTCATAAT 1146
 Qy 379 GlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGly 398
 Db 1147 CATTAATTAATTAATTAATGAACAGTCACTCTTTAAGGGATTAGATCGGATCTTGT 1206
 Qy 399 ValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGln 418
 Db 1207 TTTAGC-----CTTAAAAATAGGAATATCGAAGAGAGTGGTACTCAGCGCTATTACCTAAT 1263
 Qy 419 ProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn 438
 Db 1264 GTAAGTACTATTAGTAAGCTATA-----TTCTTACAGAAAGAAATTAAT 1308
 Qy 439 TrpAspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArg 458
 Db 1309 ATCAACAATTTACTTTTAAATACGGGCTATCGTTTGAAGAAGAAATTAATGATGCTT--- 1365

Db 748 AAGGGCGCGACTAC----- 762
 QY 231 ArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPheTyrArg 250
 Db 763 ---CGGAGCGGCAACACGACACGATATCGACGAGCTGCTCTCAGACCCATGGGAG 819
 QY 251 ArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGlnTyrAsp 270
 Db 820 CTCACCGACAGCGACAG-----TTGGCGGCCAACTTCCACTACTAC--- 861
 QY 271 AspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeu 290
 Db 862 GAGCCCTACCGCGATATG----- 879
 QY 291 TyrProHisLeuLeuThrGluGlnAspValAspTyrAspAsnProGlyLeuSerCysGly 310
 Db 880 ---CCGGCGGCTGACCCAGCGCGAGTACGAC---GACGATCCCTTCCAGTCGTA--- 930
 QY 311 PheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeu 330
 Db 931 -----CGCGACTGGGCAATATTC 948
 QY 331 ArgAsnLysArg----- 334
 Db 949 CTGGTGGCGGCAAGGACTTCTCGTGAAGTACACCCGCGAGTCCGACGACCTCACCAG 1008
 QY 335 TyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHis 354
 Db 1009 TTGAGGTGCTCACTACTACAGCAGACAGTTTCGCGCGCAGCAGCATCCGCGCGCAAC 1068
 QY 355 LeuAsnArg-----AsnAspTyrHisAspGlnLysAlaGlyAspAla 369
 Db 1069 CTAGGACCATACCTCGTACCGCGGACTACATGTGTCGCGTGAGCGCGGCTC 1128
 QY 370 ValGluAsnPhePheAsnAsnGlnTrpGlnAsnAlaArgIleGluLeuArgHisGlnPro 389
 Db 1129 TCGGGATCTTCTGCGCGCCCGACCC-----CAGGAG 1164
 QY 390 IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
 Db 1165 GTCGG-----ATCGGCTACCGTACTCTGAAGAA----- 1194
 QY 410 SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsn----- 424
 Db 1195 ---GCGATGACGAGCGCGCGCAGCTGCGCCCTGTGTCACATGTGCGCAGCGTGCCT 1251
 QY 425 -----LysValGlnHisTyrSerPhe 431
 Db 1252 CCGGCTCGGACGCCCATACCTACAGACCGTACCGCGGTACCGGCGCAGCGCTTC 1311
 QY 432 PheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGlyGlyValArgValGlu 451
 Db 1312 TACATGACGACAGATCGATGCGGACATGACGACCATACCCCGCGCATCCGCTTCGAG 1371
 QY 452 LysGlnLysAlaSerIleArgTyrAspLysAlaLeulle----- 464
 Db 1372 AAGATGACAGCGACTGGCGC---GACCGCGCGTGTCTCGGCTCGGCAACGCGCGGTG 1428
 QY 465 -----AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArg 481
 Db 1429 CAGGAGAAAGCGCAGCAGGACTACACGACCGCTGCGCGCTG----- 1476
 QY 482 GlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeu 501
 Db 1477 -----AGCTGTGTATCACTTACGCGCAGGTGG----- 1506
 QY 502 SerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGly 521
 Db 1507 AAGCTGTTCGCAACTACTCGCAATCGTTCGCGCAGCTGAG---TACTTCCAGCTCGGC 1563
 QY 522 LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSer 541
 Db 1564 CAGGCGCGCAGCGCACGATACCGCGCGCGC-----CTGGAGCGCGAGAGGCC 1614

QY 542 AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeu 561
 Db 1615 AAGACCTACGAACCTGGGCGCGCTACGACACGCGAGCTGGGCGCGGAGATCACGCTG 1674
 QY 562 TyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyPro 581
 Db 1675 TTC----- 1677
 QY 582 LysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAsp 601
 Db 1678 ---TACATCGACTTCGACGAGAGCTCAGTACGTACG---AACGACGTCGCGCTGAC 1731
 QY 602 PheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyVal 621
 Db 1732 AACCTCGCGCGC-----ACCAAGCACCAGGCGATCGAGACT 1767
 QY 622 SerGlyAspTyr-----ValArgGly 628
 Db 1768 TCCGGCAGCTATGACTTCGCGCTCTCGATCCGCGCTGGACGCGCTGTCGCTACGCG 1827
 QY 629 ArgLeuLys-----AsnLeuProSerLeuProGlyArgGlu 640
 Db 1828 AGCCTACCTATACCCGCGCCACCTACGAGGCGCATTCCTCTTCAAGGGCGCGAC 1887
 QY 641 AspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValPro 660
 Db 1888 CTGCGCTCTACTCGCGC-----CAGTGGCC 1914
 QY 661 AlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeu 680
 Db 1915 ACCGCGGGTGGCTACGAGGTC-----GACCTGTGACCTACACCTC 1959
 QY 681 AspTyrTyrArgValPheAlaGlnAsn-----LysLeuAlaArgTyrGluThrArgThr 698
 Db 1960 GAC-----GCTTCCGCGCTAGTGCATGCGCGCTGCGCGCGC-GCAGGGCTGAGCACCGCAG 2009
 QY 699 ProGly-----HisMetLeuAsnLeuGlyAlaAsnTyrArgArg 712
 Db 2010 CAGGCGCACTTCCACCACTACATCCGAGCCCGCGCGCGCGCGCAGTACGCGCA 2069
 QY 713 AsnThrArg-TyrGlyGluTrpAsnTrpTyrValLysAlaAsp----- 726
 Db 2070 CATTCGCGCTAGTGCATGCGCGCTGCGCTACGACTTCGCGCGCGCGCGCTC 2129
 QY 727 -----AsnLeuAsnGlnSerValTyrAlaHis 737
 Db 2130 GAACCTGAAGTTCGAGTTCGCGGGTGAAGAACCTGTTCGACAGCAGTATTTACCCCGCTC 2189
 QY 737 rSer 738
 Db 2190 CAGC 2193
 RESULT 6
 US-09-252-991A-10440
 ; Sequence 10440, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 10440
 ; LENGTH: 2370
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-10440

Alignment Scores:	
Pred. No.:	9.03e-15
Score:	239.50
Length:	2370
Percent Similarity:	Matches: 188
Best Local Similarity:	Conservative: 118
Query Match:	Mismatches: 294
DB:	Indels: 295
	Gaps: 42

US-09-936-377-2 (1-758) x US-09-252-991A-10440 (1-2370)

Qy	6	LeuLysProIleValLeuSerIleLeuIleAsn-----	17
Db	100	: :---TGCGTGGCCCTGCTGTTTGCATGACCCCTTCCTGCGCCCTGGCGGGAAAC	156
Qy	18	-----ThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeu	33
Db	157	GCCATCCCGCTGACCCCGACCACTACCGCCACCCCGTACCGAGCAGGACGTGGATTGCG	216
Qy	34	GluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThr	53
Db	217	GTGCCAAGACCGCTCAGCGTGGGACCGCCGCAACAACTGGAC-----	258
Qy	54	SerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeu	73
Db	259	-----CGGCAGAACGTCACACACATC	279
Qy	74	GlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSer	93
Db	280	AAGGAACCTGGTGCCTACGAACCGGAGTCTTCGGTGGCGGCGCCGCCACCGCT	333
Qy	94	AlaProVal-----IleArgGlyGlnThrGlyArgArgIleLysValLeuAsn	109
Db	334	GCCGGATCACCGGCTACACATCCGGGCATCGAGGAAACCGGATCTTACGCAGATC	393
Qy	110	HisHisGlyGluThrGlyAsp-----MetAlaAspPheSerProAspHisAlaIle	126
Db	394	GACGGGGTCGAACCTGCCCAACGACTCTTCACGGCGCCCTACGGCGAGACCCACCGCAAC	453
Qy	127	MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeu	146
Db	454	TACGTGATCCGGACATCGTAAGCGGTGGAAATCCTTCGGCGCCCGCCCTCGGCGCTA	513
Qy	147	TyrSerSerGlyAsnValAlaGlyLeuValAsp-----	157
Db	514	TACGCAGCAACGCCATCGCGCGCGGTGAGCTACTTCACCTCGACCGCTGGGCATC	573
Qy	158	ValAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGly	177
Db	574	ATCAAGAGCGCAAG-----GACGTGCGC	597
Qy	178	LeuArgLeuSerSerGly-----AsnLeuGluLysLeuThrSerGlyGlyIle	193
Db	598	GCCCGGTGAAGCGCGGTACGAGTCGGCGCAGCCACTCTGTGCTGACCTCGGCCCCCGTC	657
Qy	194	AsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu-----Tyr	209
Db	658	---GCCCGCCCGCGCAGACTTC-----GACGGCTCTGCTCATTTATGGCTAC	702
Qy	210	ArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSer	229
Db	703	CGCCAGGG-----CCACGAGACCGAATCCACGGCGG-----	734
Qy	230	ProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPheTyr	249
Db	735	CCACGGCGG-----CACCGGGCTCTCGCG-----	758
Qy	250	ArgArgThrTyrSerAspArgArg-AspGlnTyrGlyLeuProAlaHisSerHisGluTyr	269
Db	759	CAGCGAAGCAACCCGAGACGCCGACGATACAGCCTG-----	798
Qy	269	rAspAspCysHisAlaAspIleLeuTrpGlnLysSer-----LeuIle	283


```
QY 184 AsnLeuGluLys-----LeuThrSer 190
Db 517 TTATCGGAAAGGATGTTCTGTATTTGCGCAAAATGACAAATTCGATGTTCTTATTAGT 576
QY 191 Gly-----GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGlu 206
Db 577 GGTTCCTATATANTATCGGATAATTTACGCACTGGTAAAGGCAACAGCTAAATAATACC 636
QY 207 GlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg-----Asn 222
Db 637 GCC---TATAACAGTTTGGGGCTTAGCAAAATTCGGTTGCGCAAAATTAATGATCGCAAC 693
QY 223 LeuLysArgLeuProAspSerProArgArgPheAlaHisGlnHisArgAlaValLeu 242
Db 694 CGCGTGAATATATCCACCGCAAACTCGTTTAA---AACAAACAGCACCAAGCAATAATG 751
QY 243 GlyTyrArgLys-----ArgPheTyrArgTyrArgTyrSerAsp 255
Db 752 AGGTGGAACGAACTTACTAATGAACAAATTCAGATCAAAATCAAAAGTTCCACGGAC 811
QY 256 ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAsp----- 271
Db 812 AAAAAGACGATCTTCTCTCTCAACACACCAACCATCACCATCAGAAAGATCAGAGTTT 871
QY 272 -----CysHisAlaAspIleIle-----TrpGlnLysSerLeu 282
Db 872 ACTCTAAGTCGAACACGTTTAGTGTAGTGTACGTTATTTAAGTATCAACAAATTCCTG 931
QY 283 IleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyr 302
Db 932 ATCAAGCACGGTATTAAAC-TATTAT-----TTAAGCCA----- 966
QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 322
Db 967 GATAATCCTTATCTA-----AATAGCATATCGCACTGTAT 1002
QY 323 AsnGlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLys 342
Db 1003 AACAAATAAACTATTGAGAAGACAGCGTAAAGTCAGTGGTGTGAAAGATCAGACTAAA 1062
QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
Db 1063 TTGACTACCGGAGT-----ATAAATTACGTAATCTCTCGAATTATCT 1107
QY 359 -----AspTyrHisHisAspGluLysAlaGlyAspAlaVal 370
Db 1108 CACATTTCTTGTATGCGGTGATATATGCGAGATAAAATCCGTACCGAA----- 1161
QY 371 GluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle 390
Db 1162 -----CGAGGCACAAACGGTAGCGATCGAGTTTCGAGCGGACCCCTAT 1206
QY 391 GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSer 410
Db 1207 AATGCGGAATTCAAACACTACAGCGGTT---TATTTA----- 1239
QY 411 AlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer 430
Db 1240 -----ATCGCCCATATTCG 1254
QY 431 PhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgVal 450
Db 1255 CTATTTGG----- 1263
QY 451 GluLys-----GlnLysAlaSerIleArgTyrAsp-----LysAla 462
Db 1264 GAAAAATTTGCTAGTTTCGCCAAGGTACGTATGACCACCTACGTACCTCAAGTAAAC 1323
QY 463 LeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln 482
Db 1324 GTAAATACAGGATAATCAITTA----- 1347
```

```
QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTyrPyrPheThrProGlnHisLysLeuSer 502
Db 1348 -----TCTCTGCCCAAAAATAACTTGGATAGTACC-----AATGGTGTAGAT 1392
QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
Db 1393 TTTACTGCCAAATATAATAGAGCTTTCCGAGCACCATCTATGCAAGACGCGATTGTGAGT 1452
QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
Db 1453 GGTGCTCACITTCGGGCAATATCTCTAGGGCTAGATCACATCATAGATTGTAGCAAT 1512
QY 534 LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeu----- 546
Db 1513 CCAAAATTTGGCCCTGAACAGCGAAAAATAAGAAATTAACCGCAATCTACATTTTGTAT 1572
QY 547 AlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
Db 1573 AGTCTGTTTAAACAGGCGATAAATTCAAAATTTGAAGCGCATTTATTTCCGTAATGATGTG 1632
QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
Db 1633 AAAGATTTTATTAACTTAAAAATATTTATGATGCAAGACAGTGCAGTGCAGTGCA 1692
QY 580 GlyProLysSer-----IleGluAspAspSerGluMetLysLeuValArgTyr 595
Db 1693 AATCCAAATACAAATGGAGCATTTGTGCCAAAAAATTCAGATATCAA----- 1740
QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
Db 1741 AACATAACTAATGCCGTTTAACGGGTATTGAATTCGAGCTCAATACCAA---ACAGAA 1797
QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 1798 CGTTTAAAGCTA-----TTTACTAATACTATGGCAGC 1827
QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 1828 ACCAAAGGTAAAGATAAGATAGTGGC-----GAAAGCT 1860
QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
Db 1861 TTATCAACATTCGCCGACAGCAAAATCGCGTAGGGGTAAATTTATGCTTTAGTAAAAAGAC 1920
QY 675 -----ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
Db 1921 AAATTCACGTGGGAGCGACAGTAAACCATACGCT-----GCTCAACGCCGAGTGCCT 1974
QY 693 ArgTyrGluThrArgThr---ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
Db 1975 AAAGATCATAGTGTACCTACCCCAAGTTATATATCTACCGCATCTTCGTGCTACCTATGCA 2034
QY 712 ArgAsnThrArgTyrGlyGluTyr 719
Db 2035 -----CCATTAAAGCGCAATGG 2052
```

RESULT 8

US-09-557-884-1

; Sequence 1, Application US/09557884

; Patent No. 6506581

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: The Nucleotide sequence of

; the Haemophilus influenzae Rd Genome, Fragments

; Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: MD

; COUNTRY: USA

; ZIP: 20850

```

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

```

```
QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeuSer 502
Db 293031 -----TCTCCTGCCACAAAATTAACTTGGATAGTACC-----AAATGGTTAGAT 293075
QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
Db 293076 TTTACTGCCAAATATATGAAGCTTTCCGAGCACCATTCTATGCAAGACGATTTGTGAGT 293135
QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
Db 293136 GGTGCTCACTTTGGGCAAAATACTCTAGGGCTAGATCACATCAATAGATTGTAGCAAT 293195
QY 534 LysHisLeuAsnLysGlyArgSerAsnAsnIleGluLeu----- 546
Db 293196 CCAAAATTGGCCCTGAAACAGCGAAAATAAGAAATTTACCGCAATCTACATTTTGTAT 293255
QY 547 AlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
Db 293256 AGTCTGTTTAAACAAGCGGATAAATTCAAAATTGAAGCGACTTATTTCCGTAATGATGTG 293315
QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
Db 293316 AAGATTGTTTAACTTAAATAATTTAATGATGCAAAAGCAAGTGCAGGTGCA 293375
QY 580 GlyProLysSer-----IleGluAspSerGluMetLysLeuValArgTyr 595
Db 293376 AATCCAAATACAAATGAGCATTTGTCGCAAAAATTTCCAGTATCAA----- 293423
QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyIleTyrPheLysProThrPro 615
Db 293424 AACATAACTAATGCCGTTTAAAGCGGTATTGAAATTGCAAGTCAATACCAA---ACAGAA 293480
QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 293481 CGTTTAACGCTA-----TTTACTAACTATGCGAGC 293510
QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 293511 ACCAAAGGTAAGATAAAGATAGTGC----- 293543
QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
Db 293544 TTATCAACATTTCCGCAACAAATTCGGGTAGGTAATATGTTTAGTAAAGAC 293603
QY 675 -----ArgIleAspAlaAsnLeuAspTyrArgValPheAlaGlnAsnLysLeuAla 692
Db 293604 AATTCAAGTGGGAGCGACAGTAACCCATTACGCT-----GCTCAACGCGAGTGCCT 293657
QY 693 ArgTyrGluThrArgThr-----ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
Db 293658 AAGATCATAGTGTACCTACCCCAAGTTATATATCTGACCGATCTTCGTGTACCTATGCA 293717
QY 712 ArgAsnThrArgTyrGlyGluTyr 719
Db 293718 -----CCATTAAAGCGCAATGG 293735
```

RESULT 9

US-09-643-990A-1

; Sequence 1, Application US/09643990A

; Patent No. 6528289

GENERAL INFORMATION:

; APPLICANT: Robert D. Fleischmann

; Mark D. Adams

; Owen White

; Hamilton O. Smith

; J. Craig Venter

; TITLE OF INVENTION: The Nucleotide sequence of

; the Haemophilus influenzae Rd Genome, Fragments

; Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

QY 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLys 166

```
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
```

Alignment Scores:

```
Pred. No.: 4,79e-09 Length: 1830121
Score: 230.00 Matches: 171
Percent Similarity: 36.63% Conservative: 125
Best Local Similarity: 21.16% Mismatches: 285
Query Match: 5.70% Indels: 229
DB: 4 Gaps: 41
```

US-09-936-377-2 (1-758) x US-09-643-990A-1 (1-1830121)

```
QY 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSer 30
Db 291699 CTTTCCTTCGCAATTACACACC-----TAGTCACAGCAATCGCTAGCGCAATCC 291752
QY 31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
Db 291753 GTTGAATTAGACTTATCAACGTTATTGCGACACGAGAYCCA-----AGTAGGTTT 291803
QY 51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
Db 291804 GCTTATACCCGAGAAAAACAATCTAAA-----GATAGTCTTCTTTCTAAGCAAGCG 291854
QY 71 ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGly 90
Db 291855 ACTAGTGTTCAGATGCGTTTAGAGACATTCCTCCCAATGTTGATTTAGA-----GCGCGT 291908
QY 91 Gly-----AlaSerAlaProValIleArgGlyGlnThrGlyArgArg---IleLys 106
Db 291909 TCGAGAAGCATTCGCTCAAAACCTAATATTCGAGGGTTAAGTGATAATCGTGTGTGCAA 291968
QY 107 ValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
Db 291969 GTCATGTAGCGCTGAGCAAAATTTGATTATTAGCACATAGAGGT-----TCTTAT 292019
QY 127 MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeu 146
Db 292020 TTTCTTCCAATGTCTACTCATCCAGAAATTGAGTAATCAAGAGCAAGTAGTACCTCTTA 292079
QY 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLys 166
```


, PRIOR APPLICATION NUMBER: US 60/074,788
 , PRIOR FILING DATE: 1998-02-18
 , PRIOR APPLICATION NUMBER: US 60/094,190
 , PRIOR FILING DATE: 1998-07-27
 , NUMBER OF SEQ ID NOS: 33142
 , SEQ ID NO 1375
 , LENGTH: 1827
 , TYPE: DNA
 , ORGANISM: Pseudomonas aeruginosa
 , US-09-252-991A-1375

Alignment Scores:

Alignment Scores:			
Pred. No.:	2.16e-13	Length:	1827
Score:	224.50	Matches:	150
Percent Similarity:	33.5%	Conservative:	86
Best Local Similarity:	21.3%	Mismatches:	239
Query Match:	5.56%	Indels:	229
DR:	4	Gaps:	32

US-09-936-377-2 (1-758) x US-09-252-991A-1375 (1-1827)

Qy		4	T	HrThrLeuLysProIleValLeuSerIleLeuLeuleAsnThrProLeuLeuAlaGln	23
Db		155	A	GTTCTGTTCGGCGCTCCGCCATCCGCTGGTGTGCATTACAGCAGGCGGGCGA	214
Qy		24	A	LlHisGluThrGlucInSerValGlyLeuGluThrValThrValValGlyLysSerArg	43
Db		215	G	CCAAGAGAAACCACAAGAGCTGGC-----ACGGTCACCGTGTCGCGACTGGCTG	268
Qy		44	P	rArAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysfillelle---Ser	62
Db		269	G	GGGAAGACCGACAGCGCGTGTGCAGAACCATCCCGTGC CGCACGCTGGTGC	328
Qy		63	G	LlYasPThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly	82
Db		329	C	GGAAATGCTCAGAGCGGTGGCAGAAAGTCGCGAGCTGCTGCCAAGTGCCTGCGC	398
Qy		83	I	lHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal-----Ile	97
Db		389	G	TCCAAGTGCAGACAAACAACGGTAACCGCGCGCAGCATCTCCCTGAACGTGGGGTG	448
Qy		98	A	rGlyGlnThrGlyArgArgIleLysValLeuAsnHisGlyGluThrGlyAspMet	117
Db		449	C	GTGGCTGACTTCAGC-----ThrAlaLeuSer-----GlnGlnValGlu	456
Qy		118	A	laaspPheSerProAspHisAlaileMetValasp-----	129
Db		467	-	-CTGTGCGCGCGCTCGACGCTGATGATCGACGGCGTGC CGCGCGGTGGCGCCC	520
Qy		130	-	------ThrAlaLeuSer-----GlnGlnValGlu	137
Db		521	T	ACGGCAGCGCGAGCTGCTGATGATCGCGCTGTCCATCGGCAACCTGGAGAGCATCGAC	580
Qy		138	I	leLeuArgGlyProValThrLeuLeuTyrrSerSerglyAsnValAlaGlyLeuValasp	157
Db		581	G	TCTGTCGCGCGCGCGCTGCTGCTGCGCTACGGCCGCGCAGAACGTGGCGGGTGATCAAC	640
Qy		158	V	alAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGly	177
Db		641	T	TTCGTGACCCGGGGGATTCCGGAGASGTC-----TCGGCGCAATAATCGGC	685
Qy		178	L	euaArgLeu-----SerSerglyAsnLeuGluLysLeuThr-----Ser	190
Db		686	A	CCACCATCGAGCATGCGCGCCACGCGCGCTGGAAGAAGCTCAACACGCGCTTCTCTCGGC	745
Qy		191	G	lyGlylleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArg	210
Db		746	G	GGAGCCGCCAGAACGGCCCTGGGC-----GTGGCGCTGCTGTATTCTCGGGGTG-	793
Qy		211	L	ysserGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerPro	230
Db		794	N	AGGGCGCGGACTAC-----	808

QY		231	ArgArgPheAlaasnGlyGlnHisargAlaValLeuGlyTyrArgLysArgPheTyrArg	250
Dd		809	---CGGACGGCAACAACAGCAGATATCGACGCTGCTCTCAAGCCCATTTGGCAG	865
QY		251	ArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGlnTyrAsp	270
Dd		866	CTCACCCAGACGCCAGCAG-----TTGGCGGCCAACACTTCCTACTACTAC---	907
QY		271	AspCysHisAlaAspIleIleTrpGlnLysSerLeulleAasnLysArgTyrLeuGlnLeu	290
Dd		908	GAGCGCTACGCCGATATG-----	925
QY		291	TyrProHisIleuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGly	310
Dd		926	---CCGGCGGCCTGACCAGCGCAGTAGCAC--GACGATCCCTTCCAGTCGTA---	976
QY		311	PheHisAspAspAspAlaHisAlaHisAsnGlyLysProTrpIleAspLeu	330
Dd		977	-----CGCAGCTGGGACAATTTC	994
QY		331	ArgAsnLysArg-----	334
Dd		995	CGTGTCGGCGCAAGACTTCTCGTGAAGTACACCGCGCAGCTCGACGACCTCACCCAG	1054
QY		335	TyrGluLeuArgAlaGlnTrpLysGlnProPheProGlyPheGluAlaLeuArgValHis	354
Dd		1055	TTGAGGTGCTCACTACTACAGCGCAGATTCCGCGGCAGCAGCATCGCGCGCCAAC	1114
QY		355	LeuAsnArg-----AsnAspTyrHisHisAspGlulylsAlaGlyAspAla	369
Dd		1115	CTCAGGACCATCACTCGTACCGCGCGACTACCATGTGTTCGCGGTGGAGCCGCGGTC	1174
QY		370	ValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro	389
Dd		1175	TCGCGGATCTTCTTCGCGCGCCGACCAC-----CAGGAG	1210
QY		390	IleGlyArgLeuLysGlySerTrpGlyValGlnTyrIleuGlyGlnLysSerSerAlaLeu	409
Dd		1211	GTCCGC-----ATCGGCTACCGTACTCTGAAGA--	1240
QY		410	SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsn-----	424
Dd		1241	---CGGATGACAGCGCGCCAGCCAGCTGGCCCTGTGTCAACATGTGCCGACGGTCGT	1297
QY		425	-----LysValGlnHisTyrSerPhe	431
Dd		1298	CCCGGCTCGAGCGCCATACTACAGGACCGTACCGCGGTACCGAGCGACGCGCTTC	1357
QY		432	PheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgValGlu	451
Dd		1358	TACATCGACGACAGATCGATGTCCGCAACTGGACCATCACCCCGCGCATCCCGTTCGAG	1417
QY		452	LysGlnLysAlaSerIleArgTyrAspLysAlaLeulle-----	464
Dd		1418	AAGATCAGACGAGCTGGCGC---GACCGCCCGGTGCTCGCCCTGAACGCAACCGCGTG	1474
QY		465	-----AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArg	481
Dd		1475	CAGGAGAAGAGCCGACCAAGGACTACAAAGAACCGCTGCGCGGTG-----	1522
QY		482	GlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeu	501
Dd		1523	-----AGCCTGATGTATCACTTGAGCGACGAGTGG-----	1552
QY		502	SerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGly	521
Dd		1553	AAGCTGTTCGCAACTACTCGAATCTGTCCGACGACCTGCAG-----TACTTCCAGCTCGGC	1609
QY		522	LysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLysAsnLysGluArgSer	541
Dd		1610	CGGCGCGCGCAACGATACCGCCCGCGGC-----CTGAGCGCGGAAGGCC	1660
QY		542	AsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeu	561

1039 AATTGGCGTACACCGCAAAAGATTGGCTTTGGCTTGTCTGAAATAAAGTGGATGCCAACGCTAT 1098
 321 AlaHisasnGlyLysProThrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGlu 340
 1099 GTGCT---GGAAAAAAGCGTATTTCGCGCGATGACAAAGATAACCGCTACGCGAGCAA 1154
 341 TrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
 1155 TGTAAAGAGCGCCCAACCA-----TACCGAATCGCCACTCGGGG 1193
 361 HisHisaspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsn 380
 1194 CAT----- 1196
 381 AlaArgIleGluLeuArgHisGlnPro-IleGlyArgLeuLysGlySerTrpGlyValGlu 400
 1197 ---GAACCTCAACTTCACAGCGCGCTTGCAGAACAAACCTGTTGAAATACGCGCATCAA 1253
 400 nTyrLeuGlyGlnLys-----SerSe 407
 1254 CTACCGCATCAGGAATCAACCGCAAGCGTATTGAATTCACATTTAAATTTGAAGA 1313
 407 rAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlu 427
 1314 TAAAAAGATGCAACTGAGGAAGATAAATGGAAGAACCGTGCAAAATGAAAAAATTGCCAA 1373
 427 nHisTyrSerPhePheGlyValGluGlnAlaAsnTrp----- 439
 1374 AGCTTACGCTGTACCAACCGCAACCGATACCGCGCGTATATCGAAGCCATTCA 1433
 440 ---AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIl 457
 1434 CGAGATTGACGGCTTTACCTGTACCGCGCGCTCGTTTACGACCGCTTCAAGGTGAAAAAC 1493
 457 eArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys---GlnProLeuProAs 476
 1494 CCAC-----GACGG-CAAAACCGTTTCAAGCAGAGCGCTCAACCGCA 1534
 476 pLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheT 496
 1535 GTTT-----CGCGGTGATTGGCAACCGCGCGCGCAACACACTGGAGCTTCA 1576
 496 hrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnG 516
 1577 GCGGAGGCAACAATAC-----GCCAGCGGAGCGCGCGCTG-----TATGACG 1621
 516 luLeuTyrAlaHisGlyLysHis-----ValAlaThrAsnThrPheGluValG 532
 1622 CGCTGCAAAACCGCAACCGCATCTCTCGATTGCGGCGGCGAC 1671
 532 lyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluG 552
 1672 -----AAACCGGACCGCGCGGCAATACCGAAATCGGTTCACTACATCAACG 1717
 552 lyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg----- 563
 1718 ACGGACGCTTTCGCAAAACCGCACTACTTCGCGCAGACCATCAAAAGCGGCTTGCCA 1777
 564 ---AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
 1778 ATCCGCAAAACCGCAACCGCTTTTACGATACCCATCCCTCAAAAACTGTTGAGCGGCA 1837
 578 ----- 578
 1838 AAAACACCGTTACGAATTGGCGGCTCTTACCGCAGCGCGGCTGACCGCAAGTCG 1897
 579 -----ArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrA 596
 1898 GGTAAAGCGCGCAACCGCGCTTTTACGATACCCATCCCTCAAAAACTGTTGAGCGGCA 1957
 596 snGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyGluIleTyr----- 610
 1958 ACCCGAG-----TTTGGCGCAACACCGCGCGGCACTTGGACGCGCTCCCTTG 2005

611 -----PheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgG 628
 2006 CCTACCGCTTCAAA---AACCGAATCTGGAATCGCTGGCGGAGCGTATGTT--- 2058
 628 lyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProp 648
 2059 -----CAAAAGCTACGCGTTCGATATGG 2093
 648 heIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisL 668
 2094 CGGAGGCGCAAAAGACCGGACCGCAAAATTGGAACCGTTGTACGCCACGAGTTTCGGTG 2143
 668 euLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaG 688
 2144 TGAAC-----GATGCTTCGCCCA 2161
 688 InAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyA 708
 2162 ACTGAAACCGCTGGCAAGACACGCTCAAT-----GTTAATCTTCGG 2206
 708 laAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnL 728
 2207 TTAAC-----AACG 2215
 728 euLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetG 747
 2216 TGTTCGACAAGTTCTACTATCGCACAGCAACGCTGGACCAATACCTGCGCGCGTGG 2275
 747 lyArgSerPheThrGlyGlyValAsnValLysPhe 758
 2276 GACGTGATGATCGCTGGCGTGAACACTCAAGTTTC 2310

RESULT 12
 PCT-US96-05320A-113
 ; Sequence 113, Application PC/TUS9605320A
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences
 ; APPLICANT: 9410 Key West Avenue
 ; APPLICANT: Rockville, MD 20850
 ; APPLICANT: United States of America
 ; APPLICANT: Johns Hopkins University
 ; APPLICANT: 720 Rutland Avenue
 ; APPLICANT: Baltimore, MD 21205
 ; APPLICANT: United States of America
 ; APPLICANT: Mark D. Adams
 ; APPLICANT: Hamilton O. Smith
 ; APPLICANT: J. Craig Venter
 ; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
 ; NUMBER OF SEQUENCES: 48
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 ; STREET: 1100 New York Avenue, Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20003-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US96/05320A
 ; FILING DATE: April 22, 1996
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/476,102
 ; FILING DATE: June 7, 1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/487,429
 ; FILING DATE: June 7, 1995

[illegible]

Db 1534 GCACCTCTTTAAATTTAGAAAAGAAATGATTTAATCAACGAGAGTGGGTGGAAATT--- 1590
Qy 546 LeuAlaLeuGlyTyr-----GluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyr 562
Db 1591 ATTACTCTCAATCATCTATTTCAGGAAATATGATCTTTTCAATTTAGAACTACTTATTTT 1650
Qy 563 ArgAsnArgPheGlyAsnTyrIleTyr-----Ala 572
Db 1651 TACAATCGCGCAAGAATGAATATTTAAACACGAGGAGTAAATGTGTAGGCAATGCT 1710
Qy 573 GlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeu 592
Db 1711 GCAGATACCAATAATAGTTTGTCTTAAATCATTTGAAAT----- 1752
Qy 593 ValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLys 612
Db 1753 ---TATCGTATTTGCCAGGTATGTATTCAAGGTGCAGAGTTAGAAGCTTATTATCAA 1809
Qy 613 ProThrProArgTyrArgIleGly---ValSerGlyAspTyrValArgGlyArgLeuLys 631
Db 1810 TCTACT-----TATTATTGTGTGAGATACATATTCTTATGTAAAGGAAACGGTAC 1863
Qy 632 AsnLeuProSerLeuPro---GlyArgGluAspAlaTyrGlyAsnArgProPheIleAla 650
Db 1864 ACCTCACCAGAAACCCATCGGGTAAACCTCAACATGG-----ATTGCT 1908
Qy 651 GlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAla 670
Db 1909 -----GAAATPCCCCCTAGAAAACCACTACTGCTTTAGGTTTCAATGTTCCTCAA 1959
Qy 671 SerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLys 690
Db 1960 TAT-----TATCTCACAGTAGGTGGCGGTGAG 1989
Qy 691 LeuAlaArgTyrGluThrArgThrPro-----GlyHisMet 703
Db 1990 TTTGTAGAGACACAAGATCGATCGCCATTATCTGGTGATCTCTAAAGCATCATCTTGTC 2049
Qy 704 LeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyr----- 719
Db 2050 TTGCCAGCTTCAAGAGATACAGCTACATAACCTATTCTCTTGGAGCCCTGCAAAA 2109
Qy 720 -----AsnTrpTyrTyrLysAlaAspAsnLeuAsnGlnSerValTyrAla 735
Db 2110 ATTAAGGTATGATGTAGATTACAGTGTGATTTATTTCAATCAGCATAC----- 2163
Qy 736 HisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThrGly---GlyVal 754
Db 2164 -----AACCCTTATCTAGCGAAATTGGCTTCTGGACAGGTAGG 2202
Qy 755 AsnValLysPhe 758
Db 2203 AATATCAAAATTT 2214

RESULT 13
US-09-252-991A-12930/c
; Sequence 12930, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12930
; LENGTH: 618

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12930
Alignment Scores:
Pred. No.: 2.82e-13 Length: 618
Score: 216.00 Matches: 64
Percent Similarity: 45.13% Conservative: 38
Best Local Similarity: 28.32% Mismatches: 90
Query Match: 5.35% Indels: 34
DB: 4 Gaps: 6

US-09-936-377-2 (1-758) x US-09-252-991A-12930 (1-618)

Qy 506 SerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValala 525
Db 616 AGCTACACCGGCGCGCACCGACCTTCTACGAGCTGTACGCCACGCTCCGACGCCGCC 557
Qy 526 ThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGlu 545
Db 556 ACCGGCACCTACGAGGTAGCGATCGCGACGACGAAGGAAAGGGGGTCTCCACCGAC 497
Qy 546 LeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArg 565
Db 496 CTGCCCCCTGCGCTCGCAACCGCGGTGCACAGGCGCGTGGGGTGTCTACAGCGCG 437
Qy 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLys----- 582
Db 436 TTCTCCAACTACATC-----GGCCTGCTCGCCAGCGGTCCCATCGCAACGAGGAAGGC 383
Qy 583 -----SerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
Db 382 GAAGTGGTCCCGCCGACGATGACGAGCGCTGCGGATACCTCTACAGCGCGTTCGC 323
Qy 600 AlaAspPheTyrGlyAlaGluGlyGlu-----IleTyrPheLysProThrPro 615
Db 322 CGGCACTTCTACGGCTCGAGGCGCAGACCGCATCCACTGTGGAAGCCCGTACGGC 263
Qy 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 262 AACTTCGACCTGGAACCTCTCCGGGACTACACCGGACCAAGAACAGGAC----- 212
Qy 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 211 -----ACCGCGCAACCG 200
Qy 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArg 675
Db 199 CTGCCACGATCCCGCTGCGCTGAAACACCGCGCTGATCTGGGAGTTG---CAGCAG 143
Qy 676 IleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGlu 695
Db 142 TGGCAGCGCGGGTGCAGCTCGAACACGCGCTCGCAGCACCGCGTCCGCGGAGGAAGA 83
Qy 696 ThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArg 715
Db 82 CTCTCCACCGACGCGCTAC-----ACCACCTCGCGCGCAGCTCGGCTCAAACTTCGAC 29
Qy 716 TyrGlyGluTyrAsnTrp 721
Db 28 CTGCGCGAGAGCGCGCTGG 11

RESULT 14
US-09-252-991A-6426
; Sequence 6426, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

239	ArgAlaValLeuGlyTrp	----	244
766	GACGTCGGCTCGCGTGGACCCCGACGAGCACCCCTGATCGAACTCACCCCGCGCAAG	825	
245	-----ArgLysArgPheTyrArgArgThrTyrSerAsp	-----ArgArgAsp	258
826	GCGCAGCGCGGACGACGTACGCGCGCGCGCATGCGCGCTCGCAGTTCAAGCGCGAG	885	
259	GlnTyrGlyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAspIleIleTrp	278	
886	AGCCTGGCGCTCGCTTCGTCAAAATCGAACGTCACGAT	924	
279	GlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGlu	298	
925	-----GTGCTGGAGAGGTCCAGCAGACAGTTTAC	954	
299	AspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHis	318	
955	-----TACAATTACCCCGAC	969	
319	AlaHisAlaHisAsnGlyLysProIleAspLeuArgAsnLysArgTyrGluLeuArg	338	
970	-----CACATCAGG	978	
339	AlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn	358	
979	GACAAGTTCGCGCTCGGCACTCCCGCA	1017	
359	AspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThr	378	
1018	GATGCCCATCGCGATGGCGTCCCGAT	1053	
379	GlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer-TrpGln	398	
1054	CACCCTCGCGGTGGCTCGCGCCACTTGGCGTGGACGACTTCAAGCTGTGTCACCG	1113	
398	yValGlnTyrLeu	411	
1114	GGTGACGCGCATGCGCAAGCAGCACCGCGCGCGGCTCCAAGTAGACATGATGACCGA	1173	
411	aThr--SerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer	430	
1174	NTANTACACCGACGCGCAGCAGTTCCTCTGGAGCAGGACGCGGTGTTCCCAACTACCG	1233	
430	rPheGlyValGluGlnAlaAsnTrp	447	
1234	GGCCTTCGCG	1287	
447	yValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGln	467	
1288	GTTCGCCCTCGGAC	1315	
467	uAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPhe	487	
1316	-----GACTACCGCGCAGACCCTGAAGTCCCG	1341	
487	eAlaLeuSerGlyAsnTrpTyrPheThrPro	497	
1342	GCACATG	1398	
498	-----GlnHisLysLeuSerLeuThrAlaSer	506	
1399	GCCACGGCTTCGTGCGTACGAGCAGCAGCACTCGCGGACTCGCCGACCACTATACGC	1458	
507	-----HisGlnLysArgLeuProSerThrGlnGluLeuTyrAla	521	
1459	CGGCTTCGCGCATGCGCAGCGCTTCCCGCACTACTCGGAGCTGTTCTCGCCCAAGCGCG	1518	
521	yLysHisValAlaThrAsnThrPheGlnValGlyAsnLysHisLeuAsnLysGluArgGse	541	
1519	GCGAAGCGGTTCGTCAACGCTTCGAC	1566	
541	rAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLe	561	

957 AACG-----GGCGCAAAACCGTACCGTACAGCGCGTGGCAAAACCGCAGC 1001
 224 sArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTr 244
 1002 TACCTCGCAAAATCGGAACAACC-TTCGGGAGCGGGACCAACCGCATCGGTGTGAGCCA 1060
 244 pArgLysArgPheTrpArg-----ArgThrTyrSerAspArgAspGlnTyrGlyLe 262
 1061 TATGAAGACCAACACCGCGGCATCCGCACTGTG-----CGTGAAGAGTTTGGCGT 1111
 262 uProAlaHisSerHisGlnTyrAspAspCysHisAlaValLeuTrpGlnLysSerLe 282
 1112 CAGCGCAAAAT-----TACCGGAT 1132
 282 uIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTy 302
 1133 AACTATTAAACCGCAAGCCCATCTACCGGAA---ACCATCAATCAACCAACCAACT 1189
 302 rAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 322
 1190 GCGGTACACCGCAAAAGATTTGGGCTTTGTGCAAAAACCTGGATGCCACGCTATGTGT 1249
 322 sAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLy 342
 1250 ----GGAAAAGAAACGCTATTCCGCGGATGCAAGATTAACGGCTACGAGCAATGTAA 1305
 342 sGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisH 362
 1306 AAGGCCCAACCA----- 1318
 362 sAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsnGlnThrGlnAsnAlaAr 382
 1319 -----TACCGCATCCGCACTCGAGTATGA 1344
 382 gIleGluLeuArgHisGlnPro-IleGlyArgLeuLysGlySerTrpGlyValGlnTyrL 402
 1345 ACTTCAACTTCAGACGCCCTTGGCGAACAACCGCTGTGAATATACGCGCATCAACTACC 1404
 402 euGlyGln-----LysSerSerAlaLeu----- 409
 1405 GCATCAGGAATCAAAACCGCAAGCGTTTGTGAATCTCGAAATTTGAAATAAAGATAAG 1464
 410 --SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisT 429
 1465 AAAAGCAACTAATGAAGAGAAAAGAAAGAACCGTGAATAATGAAATAATGGCAAGCCT 1524
 429 YrSerPheGlyValGluGlnAlaAsnTrp----- 439
 1525 ACGCCTGACCAACCGCAACAAACCGATACCGCGCGGTATATCGAAGCCATTCACGAGA 1584
 440 --AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgT 459
 1585 TTGACGCGCTTACCTGACCGCGCGGCTGCTGTACGACCGCTCAAGGTGAAACCCAC- 1643
 459 YrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys---GlnProLeuProAspLeuG 478
 1644 -----GACGG-CAAAACCGTTTCAAGCAGCAGCGCTCAACCCGAGTTT- 1684
 478 LyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheThrPro 497
 1685 -----CGCGGTGATTGGCAGCGCGGCAACACTGGAGCTTCAGCGG 1727
 498 GlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeu 517
 1728 AGCCACAACATAC-----GCCGCGCGCAGCGCGCGCTG-----TATCAGCGCTCTG 1772
 518 TyrAlaHisGlyLysHis-----ValAlaThrAsnThrPheGluValGlyAsn 533
 1773 CAACCCAGCGGCAAGCGCGCATCATCTCGATTGCGCAGCGGACG----- 1817
 534 LysHisLeuAsnLysGluArgSerAsnAlaIleGluLeuAlaLeuGlyTyrGluGlyAsp 553
 1818 -----AAAGCCGAACCGCGCGCAATACCGAAATCGGCTTCAACTACAACGACGCGC 1868

554 ArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGln 573
 1869 ACCTTTTCCGCAAC-----GGCAGCTAC---TTCGGGCGAG 1901
 574 ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuVal 593
 1902 ACCATCAAGACGCGCTTCGCAATCCGCAAAACCGCACAGCTCCGTCGCCGCTCCCGCAA 1961
 594 ArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyIleTyrPheLys 612
 1962 GCGGTCAACCGCGGTACATCAAAAACCGAGTGTACGATTTGGCGCGCTCTACCGCACC 2021
 613 ProThrProArgTyrArgIleGlyValSer----- 622
 2022 GCGCGCTGACCCGCAAAAGTGGCGGTAGCCACAGCAAAACCGCGCTTTTACGATACGCA 2081
 623 -----GlyAspTyrValArgGlyArgLeuLys 631
 2082 AAAGACAAGCTGTGAGCGGAACCTGAATTTGGCGCAACAAGT-CGGCGGCACTTGGAC 2140
 632 AsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGln 651
 2141 GGCTCCCTTGCCTACCGCTTCAAAAACCGCAATCTGGAATCGG-----CTGGCGCGG 2194
 652 AspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHis-LeuLysAlaSe 671
 2195 TCGTATGTCAAAAACCGCT-----GGTTCGATTTGGCGGAGG 2236
 671 rLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLe 691
 2237 TCAAAAAGACCGC---GACGGCAAAATTGGAA-----AACGTTGTACGCCAA- 2279
 691 uAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrAr 711
 2280 -----GGTTTCGGTGTGAACGATGTCTTCGCCAACTGGAA 2314
 711 g-----ArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLe 728
 2315 ACCGCTGGCAACAGACG-----CTCAATGTTAATCTTCGTTAACACGTT 2362
 728 uLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetG 747
 2363 GTTCGACAAGTTCTACTATCGCACAGCAACGCTGGACCAATACCTCGCGGCGGTGG 2422
 747 YrAspSerPheThrGlyValAsnValLysPhe 758
 2423 ACGTATGTACCGCTGGCGGTGAACTACAAAGTTT 2456

RESULT 16

US-08-425-843-1
 ; Sequence 1, Application US/08425843
 ; Patent No. 6020154

GENERAL INFORMATION:

; APPLICANT: Hansen, Eric J.
 ; APPLICANT: Cope, Leslie D.
 ; APPLICANT: Jarosik, Gregory P.
 ; APPLICANT: Hanson, Mark S.
 ; TITLE OF INVENTION: H. Influenzae HxNB and HxuC Genes, Proteins
 ; TITLE OF INVENTION: and Methods of Use
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210-4433
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/425,843
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: AMCY:012/PAR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 418-3000
 ; TELEFAX: (713) 789-2679
 ; TELEX: 79-0924
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4061 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; US-08-425-843-1

Alignment Scores:
 Pred. No.: 1,4e-10 Length: 4061
 Score: 203.00 Matches: 170
 Percent Similarity: 35.86% Conservative: 119
 Best Local Similarity: 21.09% Mismatches: 296
 Query Match: 5.03% Indels: 223
 DB: 3 Gaps: 43

US-09-936-377-2 (1-758) x US-08-425-843-1 (1-4061)

QY 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGlnInSer 30
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 127 CTTTCCCTGCAATTGCCAACACC-----TTAGTGACAGCAATCGCGTAGCGCAATCC 180
 QY 31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 181 GTTGAATTAGACTCTCAACGTTATTGCGACACGAGATCCA-----AGTAGGTTT 231
 QY 51 LeuHisThrSerThrAlaSerAspIleIleSerGlyAspThrIleuArgGlnIleAla 70
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 232 GCTTATACCCGAGAAACAACTCTAAA-----GATAGTCTTCTTTTCCAAAGCAAGCC 282
 QY 71 ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnIleGlyGly 90
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 283 ACTAGTGTCGACGAGCGTTAGACACATCCCAATGTTGATATTAGA-----GCGGT 336
 QY 91 Gly-----AlaSerAlaProValIleArgGlyGlnThrGlyArg-----IleLys 106
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 337 TCAAGAAGCATTGCTCAAAAACCTAATATTCGAGGTTAAGTGATAATCGTGTGTGCAA 396
 QY 107 ValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 397 GTCATTGATGCGTGAGACAAATTTTGATTATAGCACATAGAGGT-----TCTTAT 447
 QY 127 MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeu 146
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 448 TTCTCTCCAAATGTCATCATCAAGAAATGAAGTAATCAAGACCAAGTAGCTCCTTA 507
 QY 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGlyLys 166
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 508 TGGGTGATGCGGTGCAATGGGTGTTGTTGGCAATGCGTACGCCAAATGCTTTGACTTA 567
 QY 167 MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly----- 183
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 568 TTGAAAAATAAT-----GACAAATTCGAGTTAAATTCGCCAAGGTTATCAAACT 618
 QY 184 -----AsnLeuGluLys----- 187
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 619 GCTAATAATTTTCGAAAGGATGTTCTGTGTTTTCGGCAATGACAAATTCGATGTT 678
 QY 188 LeuThrSerGly-----GlyIleAsnIleGlyLeuGlyLysAsnPheValLeu 203
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :

DB 679 CTTATTAGTGGTTTCTATATAATATCGGATAATTTACGCACCTGCTAAAGCAATAAGCTG 738
 QY 204 HisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg----- 221
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 739 AATAATACCGCC---TATAAACAGTTTGGGGCTTAGCAAAATTCGGTTGGCAATAAT 795
 QY 222 -----AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg 239
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 796 GATGCTAACCGCGTGAATATCCACCGCGAAACTCGTTTAA--AACAAACAGACACCAA 853
 QY 240 AlaValLeuGlyTyrArgLysArgPhe----- 248
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 854 GCAATAATGAGGTGGAACAGCAATTTACCAATGAACAAATACAGATCAAAATCAGAGAGT 913
 QY 249 -----TyrArgArgThrTyrSerAspArgAlaGlnTyrGlyLeuProAlaHisSer 266
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 914 TCCACAAACCAACACGTTCTCCACCGAAGCCAAACCAT-----CACAAAG 961
 QY 267 HisGluTyrAspAspCys-----HisAlaAspIleIle-----TyrGln 279
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 962 AAGAGTTTTTACTCTGCGCGTGAACACACAGCTTTTGGTAGTGTCTATTATTAACGTCAAC 1021
 QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAsp 299
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1022 AAATTCCTGATCAAGACGCGTATTAACTATTAT-----TTAACGCCA----- 1065
 QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAla 319
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1066 -----GATAATCCITATCTA----- 1092
 QY 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1093 GCACGTGTATACAAATAAACTATTGAGAAAGACACGCTAAAGTCAGTGTGTGAAGAT 1152
 QY 340 GluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn--- 358
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1153 CAGACTAAATTCACACCGAGGT-----ATAAATTTAGTAATCTTCTCC 1197
 QY 359 -----AspTyrHisHisAspGlyLysAlaGly 367
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1198 GAATATCTCACATTTCTTTGTTTATGCGGGTGGATATATCGGAGATAAAATCCGTACC 1257
 QY 368 AspAlaValGluAsnPheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHis 387
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1258 GAA-----CGAGGCACAAACAATAAAGATGCGCAGTTTCGAGCG 1296
 QY 388 GlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSer 407
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1297 GACCCCTATATCGCAATTCAAACACTACAGCGGT-----TATTTA----- 1338
 QY 408 AlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGln 427
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1339 -----ATCGCC 1344
 QY 428 HisTyrSerPheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGly 447
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1345 CATATTCGCTATTGGG-----GAAAAATTCCTACTTTTCGCCAAGT 1386
 QY 448 ValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGlu 467
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1387 GTACGTTATGACCTACATACCTCA-----AGTAAACCTGTAAATACAGAGAT 1437
 QY 468 AsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPhe 487
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1438 AATCATTTA----- 1452
 QY 488 AlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis 507
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1453 GCCACAAAATTAATCTGGATAGTACCC-----AATTGGTTAGATTTTACTGCCAAATAT 1506
 QY 508 GlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAla 525
 DB |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : : |||||: : : :
 1507 AATCAAGCTTTCGAGCACCACATCTATCGAGAGCGGTTTGTGCTGGTTCGACTTGGG 1566

QY 526 Thr-----AsnThrPheGluValGlyAsnLysHisLeuAsn 537
 DB 1567 ACAAGTATTCTAGGGCGAAATGAATCATATAATTT---GTAGCAATCCAAATTTGGCG 1623
 QY 538 LysGluArgSerAsnAsnIleGluLeu-----AlaLeuGlyTyr 550
 DB 1624 CTGGAACAGCGGAAATAAAGAAATACCGCAATCTACATTTTGATAGTCTGTTTAA 1683
 QY 551 GluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIle 570
 DB 1684 CAAGCGATAAATCAAAATGAAGCGACTTATTTCCGTAATGATGTGAAGATTTTAT 1743
 QY 571 TyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGluMet 590
 DB 1744 AACTTAATAATATTAAATGATGACAAACAAATACAAATGCAAGTCAAGTGCA----- 1797
 QY 591 LysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluLeuTyr 610
 DB 1798 -----GGTGCAGGTGCA-----GGTCAAAATCCAAATCGAGCA 1830
 QY 611 PheLysProThr---ProArgTyrArg-----ileGly 620
 DB 1831 TTGTTGCCACAAAATCCAGTATCAAAACATAACTAATGCCGTTTAAGCGGTATGAA 1890
 QY 621 ValSerGlyAspTyrValArgGlyArgLeu-----LysAsnLeuProSerLeuPro 637
 DB 1891 TTCAAGAGCTCAATACCAACAGAACGTTTAAACACTGTTTACTAATATATGCGACCAACAA 1950
 QY 638 GlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaPro 657
 DB 1951 GGTAAAGATAAGATAGTGC-----GAAGCTTTATCA 1983
 QY 658 ArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp----- 674
 DB 1984 AACATTGCCCAAGCAAAATCGCGTAGGGTAAATATGCTTTAGTAAAGACAAATTC 2043
 QY 675 ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyr 694
 DB 2044 ACGGTGGGAGCAGACGTAACCCATTACGCT-----GCTCAAGCGCGAGTGCTTAAAGAT 2097
 QY 695 GluThrArgThr---ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsn 713
 DB 2098 CATAGTGTACCTACCAAGTATATATCTACCGCATCTTCGTCTACCTATGCA----- 2151
 QY 714 ThrArgTyrGlyGluTrp 719
 DB 2152 CCATTAAAGCGAATGG 2169

RESULT 17

US-09-252-991A-14350
 ; Sequence 14350, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 14350
 ; LENGTH: 1854
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-14350

Alignment Scores:

Pred. No.: 1.29e-10 Length: 1854

Score: 198.00 Matches: 157
 Percent Similarity: 34.26% Conservative: 90
 Best Local Similarity: 21.78% Mismatches: 254
 Query Match: 4.91% Indels: 220
 DB: 4 Gaps: 39

US-09-936-377-2 (1-758) x US-09-252-991A-14350 (1-1854)

QY 7 LysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGlu 26
 DB 70 GAACCGGTGAGCTTGGCGACCGAGGTGTGGCCAGCCCGCTACCGCCGAGACC----- 123
 QY 27 ThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAla 46
 DB 124 -----GCTTCCAGAGCCTGGCGCGTGTGAGGTCTCATGCCGCGAG----- 165
 QY 47 ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeu 66
 DB 166 -----GATATCGAG 174
 QY 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
 DB 175 GCGAGCCAGCGCGCGAGCTGTGGCCAGGTGTGGCCAGGTACCGCGGTGTGGTGGCG 234
 QY 87 GlnTyrGlyGly---GlyAlaSerAlaProVal---IleArgGlyGlnThrGlyArgArg 104
 DB 235 AACAAACGCGGTTCGCGCAAGAACACACCGCTGTCTTCGCGCGGACCGAGTCCGACCAT 294
 QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAsp---MetAlaAspPheSerProAsp 123
 DB 295 GTGCTGTGTGATCGAGCGGCATCAAGTCTGGCTCGGCGAGCGCTCACCGCGCTTC 354
 QY 124 HisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
 DB 355 CAGGACTTTCGCGGTGGAG-----CTGATCGAGCGCATCGAGGTGTTCGCGCGCGCGT 408
 QY 144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspVal-----Ala 159
 DB 409 TCCAGCTTGTACGGCTCGAGGCCATCGCGGGGTGTATCCAGATCTTCACCCGCGCGGC 468
 QY 160 AspGlyLysIleProGluLysMetProGluAsnGly-----ValSer 173
 DB 469 GACGCCACGGCGCGCAAGCGCTTCTTCCTCCCGGTTACGGCACCCATCAGACCTGGAG 528
 QY 174 GlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIle 193
 DB 529 GCGAGCGCGGGGTACGCGCGCGCGCGCAAC-----GGCTGGTAC 570
 QY 194 AsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys----- 211
 DB 571 AGCTTCGGGTGAGCAGC-----TTGATATCGCGCGGGATCAATACCAAGCGCGCC 621
 QY 212 ---SerGlyAspTyrAlaValProArg-----TyrArgAsnLeuLysArgLeuProAsp 228
 DB 622 GGTACTCGGGGTATGACCGCAGACGCGCTACCGCAACCTGTCCGGCAACCTCGCGC 681
 QY 229 SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPhe 248
 DB 682 GCGGCTATCGCTTCGACAAATGGCGCTG----- 708
 QY 249 TyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGlu 268
 DB 709 -----GAATCGACGCGACGCTGTCTCAGGGCCCAAGTCGACCAAC 747
 QY 269 TyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeu 288
 DB 747 ----- 747
 QY 289 GlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSer 308
 DB 748 -----GACTATGACCGAGTTTTCGGCAAC 771
 QY 309 CysGlyPheHisAspAspAlaHisAlaHisAsnGlyLys----- 325


```

Db      1693  GGCTACGCCACCTCGACCTGCGCAGCAGTAGTACCGGCTGAAACGACGAATGGCCCTGCAG 1755
QY      628   GlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGly---AsnArg 646
Db      1753  GGCCGGATGCCAACCTGTTCCGT---GCCGACTACGAAACCGCGTATGGCTACACACCAG 1809
QY      647   Pro 647
Db      1810  CCT 1812
RESULT 18
US-09-252-991A-14267/c
; Sequence 14267, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14267
; LENGTH: 2943
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14267
Alignment Scores:
Pred. No.: 2,75e-10 Length: 2943
Score: 198.00 Matches: 157
Percent Similarity: 34.26% Conservative: 90
Best Local Similarity: 21.78% Mismatches: 254
Query Match: 4.91% Indels: 220
DB: 4 Gaps: 39
US-09-936-377-2 (1-758) x US-09-252-991A-14267 (1-2943)
QY      7   LysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGlu 26
Db      2673  GAACCGGTACGCTTGGCCAGCAGGTGGTACCACCACCCCGCTACCGCCACAGACC----- 2620
QY      27   ThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProAla 46
Db      2619  -----GCTTCCACAGCCCTGGCGCGGTGAGCGTCACTACGCCGCGAG----- 2578
QY      47   ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeu 66
Db      2577  -----GATATCCGAG 2566
QY      67   ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
Db      2568  CGCAGCCAGGCGCGCAGCGTCCCGAGCTGTGTGCGCAGGTACCCGCGGTGCGCTGCGTCCGC 2503
QY      87   GlnTyrGlyGly---GlyAlaSerAlaProVal---IleArgGlyGlnThrGlyArgArg 104
Db      2508  AACAAACGGCGGTTTCGGCAAGAACACCAACACCGCTGTCTCTGCGCGCACCGAGTCCGACCAT 2444
QY      105  IleLysValLeuAsnHisHisGlyGluThrGlyAsp---MetAlaAspPheSerProAsp 123
Db      2448  GTGCTGTGTGTGATCGCAGCGCATCAAGGTTCGGCTCGGCACGCTGGCTCACCGCGTTC 2388
QY      124  HisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
Db      2388  CAGACTTGCCGGTGGAG-----CTGATCGAGCGCATCGAGGTGCTCCGGCGCGCGT 2339
QY      144  ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspVal-----Ala 159
Db      2334  TCCAGCTGTACCGCTCGAGGCCATACCGCGCGGTGTGATCCACAGATCTTCACCGCGCGCGC 2275

```

QY 160 AspGlyIleProGluLysMetProGluAsnGly-----ValSer 173
 Db 2274 GAGCGCCAGCGCCAGCGGCTTCTCTCGCGGTTACGCGACCATCAGACCCCTGGAG 2215
 QY 174 GlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyLe 193
 Db 2214 GGCAGCGCGCGGTTCAGCGCGCGCGCGCAAC-----GGCTGGTAC 2173
 QY 194 AsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys----- 211
 Db 2172 AGCTCGGGGTGAGCAGC-----TTGATACGCGGGGATCAATCCAGCGCGCC 2122
 QY 212 ---SerGlyAspTyrAlaValProArg-----TyrArgAsnLeuLysArgLeuProAsp 228
 Db 2121 GGTACTCGGGCTATGAGCCAGACCGAGCGGTCTACGCAACCTGTCGCGCAACCTGCGC 2062
 QY 229 SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe 248
 Db 2061 GCGGCTATCGCTTCGCAATGCGCTG----- 2035
 QY 249 TyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlu 268
 Db 2034 -----GAACTGACGCGCAGCGTCTCAGGCGCAAGTGCACAAC 1996
 QY 269 TyrAspAspCysHisAlaAspIleTyrGlnLysSerLeuLeuAsnLysArgTyrLeu 288
 Db 1996 ----- 1996
 QY 289 GlnLeuTyrProHisLeuLeuThrGluLysValAspTyrAspAsnProGlyLeuSer 308
 Db 1995 -----GACTATGACCAAGTTTCGCGCAAC 1972
 QY 309 CysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLys----- 325
 Db 1971 TCCGTTTCAATGCCAACCGCGCGAGCAGAACCTGTGCGCGCGCTTC 1912
 QY 326 -----ProTyr-----IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGlu 340
 Db 1911 ACTCGTTGATCCCTGCTGCTGACCTTCAGCGCGCGCGCGAGGACAGGCGCGAT 1852
 QY 341 TrpLysGlnPro-----PheProGlyPheGluAlaLeuArgValHisLeuAsn--- 356
 Db 1851 GCCTATCAGGATGCGCGTTTCTACTCGCGCTTCGATACCTCGCGCAGCGCTGCTCGT 1792
 QY 357 ArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsn 376
 Db 1791 CAGAACGACCTGACCCCTGCGCGAA-----GGCCATGTACTGACCCCTCGGCTACGACTGG 1738
 QY 377 GlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer 396
 Db 1737 CAGAGGACGAGATCAGCAGCAGCGAAGCTTCAGCGCTCGACTCGCGCTGAACAAAGGC 1678
 QY 397 TrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaVal 416
 Db 1677 TGGTTGCGCCAGTACCTCGCGCAGTACCGT----- 1648
 QY 417 LysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlu--- 435
 Db 1647 GCGCAGGATGGCACTGAGCTGCGCGCTGACGACACACGAGCTTCGCGCTGACGAC 1588
 QY 436 -----GlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGln 453
 Db 1587 ACCGCGAGCGCGCTGG-----GGCTACGCGGTGAGCGAGCGCGCTGCTTC----- 1540
 QY 454 LysAlaSerIleArgTyrAsp-----LysAlaLeuIleAspArgGluAsnTyrTyr 470
 Db 1539 -----ACGTCATACGCGCAGCGCATTCAGGCGCGCGACTTCACGAAGACTTACTAC 1486
 QY 471 LysGln---ProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
 Db 1485 CCGGACTAGCGCAATCCGCACTGCGCGAGAGCTTCGCGCAGCGCTGAGGTCGCGCTG 1426

QY 490 SerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGlu 509
 Db 1425 AGCGGT----- 1420
 QY 510 ArgLeuProSerThrGlnGluLeuTyrAlaHisGly---LysHisValAlaThrAsnThr 528
 Db 1419 -----ACGCATGCTGGGGCACTGGCGCGTGAATGCC 1387
 QY 529 PheGlu-----ValGlyAsn----- 533
 Db 1386 TTCGTATCAACAGTCGACGACTGATCGCAACGATCGCGCTCGCGCTCGGGCGCGCC 1327
 QY 534 -----LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly 549
 Db 1326 TGGGGCGAGCGCAACACATCATCGAAGCGCGCTCCGCGCGTCAAACTGGTCTCGGC 1267
 QY 550 TyrGluGlyAspArgTyrGlnTyrAsnLeu---AlaLeuTyr-----ArgAsnArg 565
 Db 1266 AGCCAGTGGTGGTGGGACTGGAGCGCAACGCGACCTTCCTCGACCCGCAAAACCGT 1207
 QY 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly-----ArgGly 580
 Db 1206 TCCGCGCGC-----GTCAACGACGCGCAACGAGCTCGCGCGCGCGCG 1165
 QY 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyVala 600
 Db 1164 CCGCGGATGTTCAACCTGAGCTGAGCGCGCTATGATGATCGCGCAACAGGTGCGCTGGC 1108
 QY 601 AspPheTyrGlyAlaGluGlyLeuTyrPheLysProThrProArgTyrArgIleGly 620
 Db 1107 AGCGTGAC---GCCGAAGCGCGCTATGATGATCGCGCAACAGGTGCGCTGGC 1051
 QY 621 -----ValSerGlyAspTyr-----ValArg 627
 Db 1050 GGTACGCGCCTCGACTGCGCGAGGATACCGGTGAACGACGAATGGCGCTGCGAG 991
 QY 628 GlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGly---AsnArg 646
 Db 990 GCGCGGATCGCAACCTGTTCGT---GCCGACTAGAAACCGCTATGGCTACACACG 934
 QY 647 Pro 647
 Db 933 CCT 931
 RESULT 19
 US-09-221-017B-970
 ; Sequence 970, Application US/09221017B
 ; Patent No. 6444799
 ; GENERAL INFORMATION:
 ; APPLICANT: ROSS, Bruce C.
 ; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1120
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FORSTER
 ; STREET: 755 PAGE MILL ROAD
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304-1018
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows
 ; SOFTWARE: FastSeq for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/221,017B
 ; FILING DATE: 23-DEC-1998
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PP1182
 ; FILING DATE: 31-DEC-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PP1546

; FILING DATE: 30-JAN-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PP2911
 ; FILING DATE: 09-APR-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/AU98/01023
 ; FILING DATE: 10-DEC-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Monroy, Gladys H
 ; REGISTRATION NUMBER: 32,430
 ; REFERENCE/DOCKET NUMBER: 27340-20021.00
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-813-5600
 ; TELEFAX: 650-494-0792
 ; TELEX: 706141
 ; INFORMATION FOR SEQ ID NO: 970:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4661 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: circular
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: UNKNOWN
 ; ORIGINAL SOURCE:
 ; ORGANISM: PORYPHYROMONAS GINGIVALIS
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1...4661
 ; US-09-221-017B-970

Alignment Scores:
 Pred. No.: 9,44e-10 Length: 4661
 Score: 196.00 Matches: 161
 Percent Similarity: 35.95% Conservative: 105
 Best Local Similarity: 21.76% Mismatches: 297
 Query Match: 4.86% Indels: 178
 DB: 37 Gaps: 37

US-09-936-377-2 (1-758) x US-09-221-017B-970 (1-4661)

QY 33 LeuGluThrValThrValValGlyLys-----SerArgProArg 45
 DB 2282 TTGGAGGAAGTCGCTACCGGTACCGGTACAGTTACCGTTCGCTGCGTCTCTGTG 2341
 QY 46 AlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleleSerGlyAspThr 65
 DB 2342 GCAACGGAAGTCCTTACCGCTAAGGACATAGCCTCTTCTCGGCTCCTACTTCCGAGGCC 2401
 QY 66 LeuArgGlnLysAlaValAlaAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla 85
 DB 2402 TTATTGCAG-----GGGCTGAGTCCGCTCTTTGACTTCGGCCCAATCTGATGGGC 2452
 QY 86 Ser-----GlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArg 103
 DB 2453 TCTTTTCATGCAGCTGAACGCGCTTAGCAGTAAGTATATCTCATCTTATCGATGTAAG 2512
 QY 104 ArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaasp-----Phe 120
 DB 2513 CGTGTG-----TACGGCGATGTAGCGCGGTGAGCCGATTTGATCGCTATT 2557
 QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGlnIleLeuArg 140
 DB 2558 TCTCCTGATCAGATC-----GACGGATCGAATCGGTGA 2593
 QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAlaAsp 160
 DB 2594 GGTGCTTCAGTTCGCTCTACGGATCCGATCCATCGCGGGGTAATCAATGTGATC--- 2650
 QY 161 GlyLysIleProGlnLysMetProGluAsnGlyValSerGlyLeuLeuGlyLeuArgLeu 180
 DB 2651 -----ACAAAAAAGATACGAATCGACTGAGTGCATATACGTCATCGCATA 2698

QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLys--- 199
 DB 2699 TCGAAGTACAAAC---GATCGGCAACCAATACTTCGCTCGATATAAACAATCGGTAAAGTTC 2755
 QY 200 -----AenPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAsp 214
 DB 2756 AGTAGCAATACCAACTATTTCTTACCATACGATCGCTGG-----CAGATAGTCCG 2809
 QY 215 TyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAla 234
 DB 2810 TTCGAAATAAAAAAGAAAA----- 2830
 QY 235 AsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSer 254
 DB 2831 ---GGATCCGCGCAACCGGTCTTG-----GAGGAAACGATATAAGAAACTTTTCGT 2878
 QY 255 AspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAla 274
 DB 2879 GCACAGGAAATCAG---GGTGTA---AGCCAATCGCTTTCTTATATGCAACTAACAAT 2932
 QY 275 AspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeu 294
 DB 2933 CTTAGCTTCAGCGAAATGTCAGTACATTAACGT-----CAGATCTTCACTCCGACT 2986
 QY 295 LeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys-----GlyPheHis 312
 DB 2987 TTTTCGAAAAAGAGGCC---TATGACAT---GGATTATCGTGTCTTTGACGCGCTTC--- 3036
 QY 313 AspAspAspAlaHisAlaHisAsnGlyLysProTyrPheAspLeuArgAsn 332
 DB 3036 ----- 3036
 QY 333 LysArgTyrGluLeuArgAlaGluTyrLysGlnProPhePro-GlyPheGluAlaLeuAr 352
 DB 3037 ACTCGGTACGAACTATCTTTCCCAATGGTCTGCATACGCTTCTTTTCGATCCGCTTA 3096
 QY 352 gValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAlaValGluAs 372
 DB 3097 CGATCGCTTCCGTTCCGATATTGTATCATGACAAGGACAGCAGTGAGAGCCTGATCAA 3156
 QY 372 n-----PhePheAsnAsnGlnThrGlnAsnAlaArgI 383
 DB 3157 CAACCAAGTTCAGACCGAGCAACCCACATCTTTCCGGGTACGCTACGCAATAAAAAACA 3216
 QY 383 eGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuG 403
 DB 3217 TCAGATCCGATAC-----ACGCGACAGGCTCGCGGTGATTTTACACTGCTTATGGCCA 3270
 QY 403 yGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMet---LeuLe 422
 DB 3271 GAAACTGACCGCGGTTCGGATATTTCGGTGAGGAATGATCTCTCCCTATAATTGAT 3330
 QY 422 uAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTyr----- 439
 DB 3331 TACCGCAAGGACAGATGCTTCCACGCTCTCTGCTTATGTACAGATGAATGGAACCGCT 3390
 QY 440 -AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTy 459
 DB 3391 CGATTGGTTCATATGACAGCGGTTCGCT----- 3421
 QY 459 rAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGly-- 478
 DB 3422 -----CTGGTACACCATCAGGAGTTCGGTACACGAATGACGCTTAAGTATCCAT 3471
 QY 479 ----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrPr 497
 DB 3472 ACTCGCAAGTATGGCGCGCTGAACCTTCGCGCTACGTATGCTAACGGCTTAAGACTCC 3531
 QY 497 oGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLe 517
 DB 3532 CACGCTGAAA-----GAGCTTTTTCACGGAACGAAC 3564
 QY 517 uTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAs 537

```

3565 CACCACTATGGTTCGCAC-----AATCTCTATCTCGCAATCGGATCTTAA 3612
QY 537 nLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTy 557
Db 3613 GCCACAGATGCGGATTAATATGCTTTGGGCTTGAGTACAAATCAAGGCCCTATCTCGTT 3672
QY 557 rAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeu----- 575
Db 3673 CAGTGCACGGTTATGACATGACACTTCGCAATCTGATCTCCTTTATGGATATACCGAC 3732
QY 576 -----AsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLe 592
Db 3733 CTCACCCGAGCAGCAAGCTCGGGAATCAAGAAACCAAGAGTATGCCACATAGAAA 3792
QY 592 uValArgTyrAsn-----GlnSe 598
Db 3793 AGCTCGAGCGCGCCCTTGATGTCCTATGTATGATGCTCTATCGGTTGGGTATCAAGTT 3852
QY 598 rGlyAlaAspPheTyrGlyAlaGluGly-----GluIleTyrPheLysPr 613
Db 3853 AGGAGCCGGATACAGCTCGTGGAGCTTAAGAAATCTCCAGACGGATGAGTGGCTGGAAG 3912
QY 613 oThrProArgTyrArgIleGlyValSerGlyAspTyrValArg-----GlyArgLeuLy 631
Db 3913 AGCTGCAGTCACTGCGCAATGTCCACCGCATGGGTTCACACTCTGGGTGAGTATAG 3972
QY 631 sAsnLeuProSerLeuProGlyArg-----GluAspAlaTyr-----GlyAsnAr 646
Db 3973 ACTTGGCGTGAAGCCCTTTTCGGCGGTATTACAGAGCGAGGTACTACAAAGACGGCAATGC 4032
QY 646 gProPheIleAlaGlnAspAspGlnAsuAlaProArgValProAlaAlaArgLeuGlyPh 666
Db 4033 TCCG-----GACTATACCTGTGGCGACTCGCCCATCGCATCGCTTCGTCG 4077
QY 666 eHisLeuLysAlaSerLeuThrAsp-----ArgIleAspAlaAsnLeuAspTyrTy 683
Db 4078 TCATTTCCGCCCATCATCTCTGGTGAAGCGCTCGGTATAGACAACTGTTGACTAGT 4137
QY 683 r-----ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
Db 4138 GGATGATCGTCTATGGGTGTCAAT-----TATGCTACCGTAACGCCGGGA 4183

```

RESULT 20

```

; Sequence 1192, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1192
; LENGTH: 2391
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-1192

```

```

Alignment Scores:
Pred. No.: 5,76e-10 Length: 2391
Score: 193.50 Matches: 170
Percent Similarity: 32.51% Conservative: 80
Best Local Similarity: 22.11% Mismatches: 234
Query Match: 4.79% Indels: 286
DB: Gaps: 37

```

```

US-09-936-377-2 (1-758) x US-09-252-991A-1192 (1-2391)
QY 76 AlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaPro 95
Db 2377 GCGCTCGACGGTATGATCGACGGCTGCGGC-----GGCGGTGGCGCCTACGCCCA 2324
QY 96 ValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGly 115
Db 2324 ----- 2324
QY 116 AspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAla-LeuSer----- 133
Db 2323 -----GCCGAGCTGTCGATGATGCCCTGTCCATCGG 2291
QY 134 -----GlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAs 151
Db 2290 CAACTGGAGAGCATCGACGTGTCGCGCGCGCTCGGTCGCTACGCGCCGCGAGAA 2231
QY 151 nValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGl 171
Db 2230 CGTGGCGGGGTGATCAACTTCGTGACCCCGGGCGATTCCGGAGAGATTCTC----- 2182
QY 171 yValSerGlyGluLeuGlyLeuArgLeu-----SerSerGlyAsnLeuGluLysLe 188
Db 2181 -----TCGCGCAATCGCACCATCGACATCGCGCCACCGCGGCTGGAGAAGCT 2126
QY 188 uThr-----SerGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHi 204
Db 2125 CAACGAGCGCTTCTCGGGGACCGCCGACACCGCGCTGGC-----GTGGCGTGTCT 2072
QY 204 sThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLy 224
Db 2071 GTATTCGGGGTG-----AAGGGCCCGACTAC----- 2044
QY 224 sArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTr 244
Db 2043 -----CGCGACGGCAACAACGACACGATATCGACGACGTCT 2006
QY 244 pArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAl 264
Db 2005 GCTCAAGACCCANTTGGCGCTCACCACACGACGACAG-----TTGGCGGC 1961
QY 264 aHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAs 284
Db 1960 CAATTCCTACTACTAC-----GACGCTTACGCGATATG----- 1927
QY 284 nLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAs 304
Db 1926 -----CCGGCGCGCTTACCGACGCGCAGTACGAC-----GACGA 1892
QY 304 nProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGl 324
Db 1891 TCCCTTCAGTCGGTA----- 1876
QY 324 yLysProTyrIleAspLeuArgAsnLysArg----- 334
Db 1875 -CGGACTGGGACAAATTCGTCGTGCGCGCAAGGACTCTCGCTGAAGTACACCGCCCA 1817
QY 335 -----TyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPh 348
Db 1816 GGTGACGACCTCACCCAGTTCAGGTGCTCACTACTACGACGACAGTTTCGCGCGCAG 1757
QY 348 eGluAlaLeuArgValHisLeuAsnArg-----AsnAspTyrHisHisAs 363
Db 1756 CAGCATCGCGCGCGCACTCAGGACCATCCTCGTCCCGCGGCGACTACCTATGTGTT 1697
QY 363 pGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIl 383
Db 1696 CGCGGTGGAGCGCGGTCTCGCGGATCTCTTCGCGCGCGCCGACCAACC----- 1648
QY 383 eGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGl 403
Db 1647 -----CAGGAGTTCGCG-----ATCGGCTACCGCTACTCTGAA 1616

```



```

Db 481 -----CtATTGAAAGCTCAGCT----- 498
QY 181 SerSerGlyAsnLeuGlnLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn 200
Db 499 -----GAAAGCTTACCTCT----- 513
QY 201 PheValLeuHisThrGluGlyLeuIleArgLysSerGlyAspTyrAlaValProArgTyr 220
Db 514 -----GAAAGCTTATCCCGTCAAGCAGGTCTTGTCTAGGTTTCATAT 558
QY 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAla 240
Db 559 GGGCGCTATT-----GACCATAATCTTGAAGCG 585
QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyr 260
Db 586 GCTGTGTGGCGATGAGAAAGAAA---TATATCCGTTTAAATGCCAACCGT----- 630
QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIle----- 277
Db 631 -----TCGAATCGAATAGTATTACAGATGGTATGCGAATACAGATTCATCAGCA 681
QY 278 TrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGlu 297
Db 682 TGGAAAGATGG----- 693
QY 298 GluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAla 317
Db 694 -----AATGTTGATGTAGCGCTTGGCTTTTACTCCAGATGAAACACAG 735
QY 318 HisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGlu--- 336
Db 736 -----TGGTCTGAAATCACCGGTGGGAAATCTCATGCT 768
QY 336 ----- 336
Db 769 GAATCGCTTTATGCTGGAGCTTCAATGATGTTCTCAGTTTGGCGGTGAAAGTTTAGGG 828
QY 337 LeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeu-----ArgValHis 354
Db 829 CTTTGGTTTGAAGAAAGAAATATCACTGATGTGATCAAGAAATTTGAAGGCGAGGTGAAC 888
QY 355 LeuAsnArgAsnAspTyrHisHisAspGluLysAla----- 366
Db 889 TATAGCTATAACGACCATATTATGATAACTTACCTTCGCTGACACCACTTAGTTGAA 948
QY 367 -----GlyAspAlaValGluAsnPhePheAsnAsn 376
Db 949 ATGAATCATGTGTGTATGACCATGTGATGTCGCAATGCAATGCAATGCAATGCAACGC 1008
QY 377 GlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySer 396
Db 1009 CGTACATTAACTCAGCTTGGCAATGACTCTGAA---TGGAAACAAATTTAGTTTGTATC 1065
QY 397 TrpGlyVal-----GlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGlu 414
Db 1066 ACAGGCATTGATTCTCAATTCAATAAACAATGCGGTAGCATGTCTATCCCAACATGCT 1125
QY 415 AlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyVal 434
Db 1126 TCAATGAATATTCATATATCGTCAAGATATGCGTTTCAATCATACGGTGTCTTTGGTGAG 1185
QY 435 GluGlnAlaAsnTyrAspAsnPheThr---LeuGluGlyGlyValArg----- 449
Db 1186 CTTGGTTATCAATGGAGCAATCTCAACAACTGGTCACAGGTGTGCGTTTATAGTCGAGTC 1245
QY 450 ---ValGluLysGlnLysAla-----SerIleArgTyrAspLys 461
Db 1246 ACAGTTGAAGATGAAGCGGTGACTCTCAAGCAAAAGCGCTTATATACTAAGCTTGAAAG 1305
QY 462 AlaIleuLysArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArg 481

```

RESULT 22

US-08-425-843-6

; Sequence 6, Application US/08425843

; Patent No. 6020154

; GENERAL INFORMATION:

; APPLICANT: Hansen, Eric J.

; APPLICANT: Cope, Leslie D.

APPLICANT: Jarosik, Gregory P.
 APPLICANT: Hanson, Mark S.
 TITLE OF INVENTION: H. Influenzae HxNB and HxuC Genes, Proteins
 TITLE OF INVENTION: and Methods of Use
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold, White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: USA
 ZIP: 77210-4433

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION NUMBER: US/08/425,843

APPLICATION DATE: Concurrently Herewith
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.

REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: AMCY.012/PAR

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000

TELEFAX: (713) 789-2679

TELEX: 79-0924

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 4651 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-425-843-6

Alignment Scores:
 Pred. No.: 3,53e-09 Length: 4651
 Score: 190.50 Matches: 163
 Percent Similarity: 35.66% Conservative: 123
 Best Local Similarity: 20.32% Mismatches: 293
 Query Match: 4.72% Indels: 225
 DB: 3 Gaps: 41

US-09-936-377-2 (1-758) x US-08-425-843-6 (1-4651)

Qy 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSer 30
 Db 742 CTTTCCCTTGCAATTACAAACACC-----TTAGTGACAGCAAAATGCCTAGCGCAATCC 795
 Qy 31 ValGlyLeuGluThrValThrValValGlySerArgProArgAlaThrSerGlyLeu 50
 Db 796 GTTGAATTAGACTCTCAACAGTTTATTCGACACAGAGATCCA-----AGTAGGTTT 846
 Qy 51 LeuHisThrSerThrAlaSerAspIleIleSerGlyAspThrLeuArgGlnIleAla 70
 Db 847 GCTTATAGCCAGAAAACAACTCTAAA-----GATAGTCTCTCTTCTTACGACAGCG 897
 Qy 71 ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnIleGly 90
 Db 898 ACTAGTGTTCGACGAGCGCTTAGAGACATTCCTCAATGTTGATGTAGTA-----GGCGGT 951
 Qy 91 Gly-----AlaSerAlaProValIleArgGlyGlnThrGlyArgArg-----IleLys 106
 Db 952 TCGAGACACATGCTCAAAAACCTTAATATCCAGGGTTAGTGATATCTGGTGTGTCAA 1011
 Qy 107 ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
 Db 1012 GTCAATTGATGGCGTGAGACAAAATTTTGTATTTAGCACATAGAGGT-----TCTTAT 1062
 Qy 127 MetValAspThrAlaLeuSerGlnGlnValGluLeuLeuArgGlyProValThrLeuLeu 146

Db 1063 TTTCTTCCCAATGTCACCTATTTCAGAAATTCAGTAATCAAGAGACCAAGTAGTCCTTA 1122
 Qy 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLys 166
 Db 1123 TGGGTAGCGGTGCTTTGGGTGGTGTGTGTCATATGCGTACGCCAAATGCTTTAGACTTA 1182
 Qy 167 MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly----- 183
 Db 1183 TTGAAAATAAT-----GACAAATTCGAGTTAAATTCGCCAAGGTTATCAACT 1233
 Qy 184 -----AsnLeuGluLys----- 187
 Db 1234 GCTAATAATTTATCGAAAAGGATGCTTCTGTATTTCGGCAAAATGACAAATTCGATGT 1293
 Qy 188 -----LeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeu 203
 Db 1294 CTTATTAGTGTCTTCTATAATAATGCGGATAATTTCACACTGGTAAAGGCAACAGCTG 1353
 Qy 204 HisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArg----- 221
 Db 1354 AATAATACAGCC--TATAAACAGTTTGGGGCTTAGCAAAATTTGGTTGCAAACTAAT 1410
 Qy 222 -----AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg 239
 Db 1411 GATCGAACCCTGTGGAATTTATCCACCGCAAACTCGTTTTTA--AACAAACAGCACCA 1468
 Qy 240 AlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgAspGln 259
 Db 1469 GCATAATGAGGTGGAAACAGCACTTACC--AATGAAAAAATTTATAGATCAAACTCAACGAG 1527
 Qy 260 Tyr-----GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAla 274
 Db 1528 TTCCACGGCTCAACACACGGTTTACACAGAGAGCAAAACCATCATCAGAAACATCAGCG 1587
 Qy 275 AspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeu 294
 Db 1588 TTTTAC-----TCTAAAGTGAAACACAGCTTTTGGCAGTGTC-----AGTTAT 1629
 Qy 295 LeuThrGluGluAspVal-----AspTyr-----Asp 303
 Db 1630 TTAACCTGATCAACAAATTCCTGATCAAGACCGGTATTTAACATAATTATTTAACGCCAGAT 1689
 Qy 304 AsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHis 323
 Db 1690 AATCCTTATCTA-----AATACCATATGCATCGCTGATTAAC 1725
 Qy 324 GlyLysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluThrLysGln 343
 Db 1726 AATAAAACTATTGAGAAAGAACAGCGTAAAGTCAGTGGTGTGAAAGATCAGACTAAATTG 1785
 Qy 344 ProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
 Db 1786 ACTACCGGAGGT-----ATAAATTACGTAACTCTTCGAAATTAATCTCAC 1830
 Qy 359 -----AspTyrHisHisAspGlyLysAlaGlyAspAlaValGlu 371
 Db 1831 ATTTCTCTTTGTTATGGGTGGATTATATCGCAGATAAAATCCGTACCGAA----- 1881
 Qy 372 AsnPheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGly 391
 Db 1882 -----CGAGCGCAACAACTAATGATGCGAAGTTTCGAGCGGACCCCTATAAT 1929
 Qy 392 ArgLeuLysGlySerTyrGluValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAla 411
 Db 1930 GCGAATTCAAACACATACAGCGCTT--TATTTA----- 1959
 Qy 412 ThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhe 431
 Db 1960 -----ATCGCCCATATTCCACTA 1977
 Qy 432 PheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgValGlu 451

ADDRESSEE: Sim & McBurney
 STREET: 6th Floor, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,890A
 FILING DATE: 07-JUN-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/175,116
 FILING DATE: 29-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/148,968
 FILING DATE: 08-NOV-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Stewart, Michael I
 REGISTRATION NUMBER: 24,973
 REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5099 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: Join(160..2121, 2152..4890)
 US-08-487-890A-4

Alignment Scores:
 Pred. No.: 4,53e-08 Length: 5099
 Score: 180.50 Matches: 197
 Percent Similarity: 33.30% Conservative: 139
 Best Local Similarity: 19.52% Mismatches: 323
 Query Match: 4.47% Indels: 351
 DB: 1 Gaps: 48

US-09-936-377-2 (1-758) x US-08-487-890A-4 (1-5099)

QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
 DB 2155 ACTAAAAACCCCTATTTCGCCCTAAGTATTATTTCTTGCTTTTAATTCATGCTATGTA 2214
 QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
 DB 2215 AAAGCAGAACTCAAGTATAAAGATACAAAGAGCTATATCATCTGAAAGTGACACT 2274
 QY 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
 DB 2275 CAAAGTACAGAGATTTCAGAACTAGAACTATCTCAGTCTGCGAATAAATAAGAGAT 2334
 QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
 DB 2335 CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATATCAAACTAGT 2382
 QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
 DB 2383 GAAAGTATCAGCGGAGAACAGATATAATATTCGTGATCTAACACGCTATGATCCAGC 2442
 QY 83 IleHisAlaSerGlnThrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr 101
 DB 2443 ATTCAGTTGTAGAACAGGCCGTGGTGCAGGTTCTGGATATCTATTCTGCTGATGGAC 2502

QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
 DB 2503 AGAAATAGAGTT----- 2514
 QY 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGly 141
 DB 2515 -----GCTTATTAGTAGATGGTTTACCTCAAAAGCAATCTTATGTAGTCCAAAGC 2565
 QY 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
 DB 2566 CCTTTAGTTGCTCGTTTCAGGATATTCTGGCACTGGTGCAATTATGAATAATGAATATGAA 2625
 QY 160 AspGlyLysIleProGlnLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179
 DB 2626 AATGTAAAGCCCGTCGAA---ATAAGCAAGGGGGAGTTCTTCTGAGTAGTGGTAATGGA 2682
 QY 180 LeuSerSerGlyAsnLeuLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly--- 198
 DB 2683 GCCTAGCTGGTTCTGTACATTTCAAAGCAAAATCAGCAGCCGATATCTTAGAAGGAGAC 2742
 QY 199 LysAsnPheValLeuHisThrGluGlyLeuTyrArg----- 210
 DB 2743 AAATCATGGGAATTCAAACTAAATAATGCTTATCAAGCAAAATAAAGGCTTTACCCAT 2802
 QY 211 -----LysSerGlyAspTyr-----AlaValProArgTyrArg 221
 DB 2803 TCTTTAGCTGTAGCTGGAACCAAGGGGATTTGACGGGTCCGCAATTTATCTCAACGA 2862
 QY 222 AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg---Ala 240
 DB 2863 AAT-----TCAATTGAAACCAAGTCCCATAAAGATGCA 2895
 QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr 260
 DB 2896 TTAAGAGCGGTACAAAGT-----TATCATCGATTATATGCCCAACCCAGAGGATCAA--- 2946
 QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTrpGlnLys 280
 DB 2947 -----TCTGCATACTTTGTGATGCAAGATAGTGTCCAAAGCCAGATGATTATACAGT 3000
 QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292
 DB 3001 TGTTTACCTTTGCCAAAGCAGCTGCGATTTATCTCTCCAAAGAGAAACCGTAAGCGTT 3060
 QY 293 -----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCy 309
 DB 3061 TCAGATTATACGGGGCTAACCGTATCAAACTCAATCAATCAATGAATGAAGCCAGTCT 3120
 QY 309 sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAs 329
 DB 3121 TGGTTTTTAAGAGGGGGTAT-CATTTTCTGAACCAACATATATTTGGTGTATTTTGA 3179
 QY 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheG 349
 DB 3180 ATTCACACACAAATAATTTGATATCCGT-----GATATGACATTTCCCGCTTAT-- 3228
 QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGlyLeuAlaGlyAspAl 369
 DB 3229 -----TTAAGATCAACAGAAAAACGGGATGATAG 3257
 QY 369 aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi 387
 DB 3258 CAGTGGCTTTTTTATCCAAAGCAAGATTTATGTCATATCAACGCTATTGAG----- 3309
 QY 387 sdlnProIleGlyArg----- 393
 DB 3310 -----GATGCCCGAGGGCTTAATCATGAGTGGGCTTTATTTTCGATGAACACCATAG 3362
 QY 393 uLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
 DB 3363 AAAACAGCGGTAGGTATTGAATATATTACGAAAAATAAGAACAAAGCGGATCATTTGA 3422
 QY 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis--- 428

Db 3423 CAAAGCAGTGTAAAGTCTAATCAACAACATCATCTACTGACAGTTATATGCAACATAC 3482
QY 429 -----TyrValArgGlyArgLeuLysAsnLe 633
Db 3483 GCATTGGCAGTCTTTATCTCAATCAAGTAAGAAATTGGCGCCCAACACAGTGTATAACCTTA 3542
QY 429 rSerPheGlyValGlu -----TyrValArgGlyArgLeuLysAsnLe 633
Db 3543 TTCACTACTCATCTGTAGAGAAATGTTTATAAGAAAAACATAATATGTTGCAATTGAA 3602
QY 436 -----GlnAlaAsnTyr 673
Db 3603 TTTAGAGAAAAAATTTCAACAAAATTTGGCTTACTCATCAAAATGCTCTTCAATCTTGGTTT 3662
QY 440 AspAsnPhenThr -----LeuGluGlyValArgValG1 451
Db 3663 TGATGACTTTACTTCAGCGCTTCAGCAATAAGATTATTTAACTGACGCTGTTACCGCTAC 3722
QY 451 uysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLy 471
Db 3723 GGCRAAGAGTATTCA -----GAGAAAGCTAATGAAACAAAGAAATGTTTCAA 3773
QY 471 sGln -----ProLeuProAspLeuGly 478
Db 3774 AAAACAACTTACTTATACCAAAACCAACAGTAGTGTGTTAGTACAGATCATTTGTGA 3833
QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAGGTAACTCTCTTAATACAGAGACTGTAAAGTGGCGTTAATTAAGAGGAAAAA 3893
QY 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu ----- 503
Db 3894 TTATTATTTCGACGACGCAATATATGCGATTTAGCGGAAATACGTTGATTTAGGTTTAGG 3953
QY 504 -----ThrAlaSerHisGlnGlu 509
Db 3954 TATTCGGTATGACGTATCTCGCACAAAGCTAATGATCACTATTAGTGTGGTAATTT 4013
QY 509 ----- 509
Db 4014 TAAAAATTTCTCTTGGATPACTGGTATTGTCATAAAACCAACGGAATGGCTTGATCTTTC 4073
QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
Db 4074 TTATCGCTTCTTACTCGGATTTAGAAATCTCTAGTTTCTGTAATGTAT -----GGTTG 4127
QY 522 shisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 GCGGTATGCTGGCAATAATAGCGAGGTTTATGTAGTAAATTTAAAGCCTGAAACATCTCG 4187
QY 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTy 562
Db 4188 TAACCAAGATTTGGTCTCGCTCTAAAGGGAATTTGCTAATATATGATCAGTCAITTT 4247
QY 562 rArgAsnArgPheGlyAsnTyrIle -----TyrAlaGlnThrLeuAsn -----AspG1 578
Db 4248 TAGTAATGCTTATCGAAATCTTATCGCTTTCGTAAGAACTTAATAAAGTGAACACTGG 4307
QY 578 yArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn -----G1 597
Db 4308 AAAAGGCAATATGATGATATCATATATGCAAAAATGCAAAATGATTTAGTTGGCGTAAATATAAC 4367
QY 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe ----- 611
Db 4368 TGCGCAATTAGATTTTAAATGTTTATGGAACGATTTCCCTACGCTGTTGATGCAACATT 4427
QY 611 ----- 611
Db 4428 TGCTTAAACCGAGTAAAGTTAAAGATCAAAAATCAATGCTGTTGGCTCGGCTAGG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624

Db 4488 CAGTTATTATTGATGGCATTGACCCAGC -----CGTTATATCATTTGGTTTAGGCTATCA 4544
QY 624 p -----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCAAGTAATATCTTGGGGAATTAATCAATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTTCTAGGAAACAGT -----GCATTGGGTAAACAAT -----TCAAGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAAAGAAACTTACTCGGCA -----TGGCAATATC ----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGATCGGTTTATATACATGCGCAATAAA ----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 4723 -----AATATTATGCTTCGATTAGGATATATAATTTATTTCAA 4760
QY 713 nThrArgTyrGlyGluTyrAsnTyrTyrValLys -----AlaAspAsnLeuAsn ----- 730
Db 4761 CTATCGCTATGTTACTTGGAAAGCGGTGCTCAACAGCACAAGGTGCGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATCTACTCGTACGACGATCA -----GGACGAAACTA 4868
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTAGAAATGAATTC 4893

RESULT 25

US-08-478-435-4
; Sequence 4, Application US/08478435
; Patent No. 5922323
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,435
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:

Qy	199	LysAsnPheValLeuHisThrGluGlyLeuTyrArg-----	210
Db	2743	AAATCATGGGGAATTCAAACATAAAAATGCCTATTCAAGCAAAAAATAAAGCGTTTACC CAT	2802
Qy	211	-----LysSerGlyAspTyr-----AlaValProArgTyrArg	221
Db	2803	TCCTTAGCTGTAGTGGAACAACAAAGGGGATTTGACGGGTGCGCAATTATTACTCAACGA	2862
Qy	222	AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg---Ala	240
Db	2863	AAT-----TCAATTGAAACCCCAAGTCCATAAGAATGCA	2895
Qy	241	ValLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyr	260
Db	2896	TTAAAGGCGTCACAAAGT-----TATCATCGATTATTCGCCAAACACAGAGGATCAA---	2946
Qy	261	GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaaspileilelrpGlnLy	280
Db	2947	-----TCTGCATACTTTGTGATGCAAGATGAGTGTCCTCCAAAGCCAGATGATTATAACAGT	3000
Qy	280	sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro-----	292
Db	3001	TGTTTACCTTTTCGCCAAACACACCTGGGATTTTATCTCTCCAAAGAGAAACCGTAAGCGTT	3060
Qy	293	-----HisLeuLeuThrGluGlueaspValAspTyrAspAsnProGlyLeuSerCy	309
Db	3061	TCAGATTATACGGGGGCTAACCGTATCAAACTTAATCCCAATGAATAATGAAAGCCAGTCT	3120
Qy	309	sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAs	329
Db	3121	TGGTTTTTAAGAGGAGGGTAT-CATTTCTGACACACATTATATTGGTGGTATTTTTGA	3179
Qy	329	pLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGl	349
Db	3180	ATTCCACACACAAAAATTTGATATCCGT-----GATATGACATTTCCCCTTAT--	3228
Qy	349	uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspal	369
Db	3229	-----TTAAGATCAACAGAAAAACGGGATGATAG	3257
Qy	369	aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi	387
Db	3258	CAGTGGCTCTTTTATCCAAAGCAAGATTATGTGTGATATCAACAGTATTGAG-----	3309
Qy	387	sGlnProIleGlyArg-----	393
Db	3310	-----GATGGCCGAGGCGTTAACTATGCAAGTGGGCTTTATTTCGATGAACACCATAG	3362
Qy	393	uLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe	413
Db	3363	AAACACGCGTGTAGGTATTGAAATATATTTACGAAAAATAAGAACACGGCGGATCATTTGA	3422
Qy	413	rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis--	428
Db	3423	CAAAAGCAGTGTAAAGTGCCTAAATCAACAAACATCATCTTGCAGTGTATATGCAACATAC	3482
Qy	429	-----TV-----	429
Db	3483	GCATTGCAGTCTTTATCTTAATCCAAAGTAAGAATTGCCGCCCAACACAGTGATAAACCTTA	3542
Qy	429	rSerPhePheGlyValGlu-----	435
Db	3543	TTCATACTATCATCTTCGTAGAGAAATGTTTTATAAGAAAAAACATAATATGTTCCAATTGAA	3602
Qy	436	-----GlnAlaasnTrp-----	439
Db	3603	TTTAGAGAAAAAATTCACAAAAATTTGGCTTACTCATCAAAATTGRCTTCAATCTTGGTTT	3662
Qy	440	-AspAsnPheThr-----LeuGluGlyGlyValArgValGl	451
Db	3663	TGATGACTTACTCTCAGCGCTTCAGCATAAAGATTATTTAACTCGACGGTGTTCGCGTAC	3722

```
QY 451 ulysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspArgGluAsnTyrTyrly 471
Db 3723 GGCAGAGAGTATTTC-----GAGAAAGCTAATGAACAAGAGAAATGTTACAA 3773
QY 471 sGln-----ProLeuProAspLeuGly----- 478
Db 3774 AAAACACCTTACTTATACCAAAACACAGTAGGTTTGTAGTACAAGATCATTTGTGA 3833
QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAAGGTAACTCTCTAATACAGAGACTCTAAAGTCGCGTTAATTAAGGGGAAAA 3893
QY 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu----- 503
Db 3894 TTATTATTTCGCAGCAGCAATAATATGCGATTAGGGAATACGTTGATTAGGTTTAGG 3953
QY 504 -----ThrAlaSerHisGlnGlu----- 509
Db 3954 TATTCGGTATGAGTATCTCGCACAAAGCTAATGAATCACTATTAGTGTGTTAAATT 4013
QY 509 ----- 509
Db 4014 TAAAAATTTCTTGGAACTCTGTATTGTCTATAAAACCAACGGAATGGCTTGATCTTTC 4073
QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
Db 4074 TTATCGCTTTCTACTGGATTAGAAATCTCTAGTTTGTGTAATGAT-----GGTGG 4127
QY 522 sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 GCGGTATGGTGGCAATAATAGCGAGGTTTATGTAGGTAAATTTAAGCCTGAACAATCTCG 4187
QY 542 nAniLeuGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTy 562
Db 4188 TAACCAAGAGTTTGGTCTCGCTCTAAAGAGGGATTTGGTAAATATTGAGATCATGTCATT 4247
QY 562 xArgAsnArgPheGlyAsnTyrIle---TyrAlaGlnThrLeuAsn-----AspG1 578
Db 4248 TAGTAATGTTATCGAAATCTTATCGCTTGTGTGAAGACTTAATAAAATGGAATGG 4307
QY 578 yArgGlyProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsn---G1 597
Db 4308 AAAGGCCAATTATGGATATCATATGCACAAATGCACAAATTAGTTGGCGTAAATATAAC 4367
QY 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
Db 4368 TGGCAATTAGATTTTAAATGTTTATGGAACGTAATCCCTACGTTGGTATGCAACATT 4427
QY 611 ----- 611
Db 4428 TGCTTATAACGAGTAAAGTTAAAGATCAAAATCAATGCTGTTGGCTCCGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTATTATTTGATGCCATTACGCCAGC---CGTTATATCATTTGGTTTGGCTATGA 4544
QY 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCCAGTAATACTTGGGGAATTAATACAAATGTTTACTCAATCAAAACCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTCTAGGAAACAGT---GCATTGGGTAAACAAT-----TCAAGGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAAATCAACAGAAAGAACTTACTCGGCA-----TGGCATATC----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGATCGGTTTATTATCGCGAAATAA----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
```

```
Db 4723 ----- 4723
QY 713 nThrArgTyrGlyGluTyrAsnTyrTyrValLys---AlaAspAsnLeuLeuAsn--- 730
Db 4761 CTATCGCTATGTTACTTGGGACGGTGGCTCAACAGACACAGAGTGGCGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATACTCGTACGAGCATCA-----GGACGAAACTA 4868
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACTTAGAAATGAATTC 4893

RESULT 26
US-08-337-483-4
; Sequence 4, Application US/08337483
; Patent No. 5922562
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(160..2121, 2152..4890)
US-08-337-483-4

Alignment Scores:
Pred. No.: 4,53e-08 Length: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
DB: 2 Gaps: 48

US-09-936-377-2 (1-758) x US-08-337-483-4 (1-5099)
```

Qy	5	ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu	20
Db	2155	ACTAAAAACCCCTATTTTCGCCTAAGTATTATTCTTGTCTTTAAATTTTCATGCTATGTA	2214
Qy	21	LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly-----	32
Db	2215	AAAGCAGAAACTCAAGTATATAAAGATACAAAGAGCTATATCATCTGAAGTCGACACT	2274
Qy	33	-----LeuGluThrValThrValValGly-----LysSerArgPro	44
Db	2275	CAAAAGTACAGAAGATTTCAGAAATTAGAACTATCTCAGTCACTGCAGAGAAAATAAGAGAT	2334
Qy	45	ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser	62
Db	2335	CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATATTCAAAACACTAGT	2382
Qy	63	GlyAspThrLeuArgGlnLeuAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly	82
Db	2383	GAAAGTATCACCCGAGAACAGTATTAATATTTCGTGATCTAAACACCTATGATCCAGGC	2442
Qy	83	IleHisAlaSerGlnThrGlyGlyGlyAlaSerAlaProVal-----IleArgGlyGlnThr	101
Db	2443	ATTTCAGTTGAGAACCAAGGCCGTGGTGCAGTTCTCGATATTTCTATTCGTGGTATGCAC	2502
Qy	102	GlyArgGlnLeuLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer	121
Db	2503	AGAATAGAGTT-----	2514
Qy	122	ProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGlnIleLeuArgGly	141
Db	2515	-----GCTTTATTAGTAGATGGTTTACCTCAAACGCAATCTTATGATAGTGCAGAAC	2565
Qy	142	ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla	159
Db	2566	CCTTTAGTTCGCTTCAGATATTTCTGGCAGTCTGGTGCATTAATGAATGAATATGAA	2625
Qy	160	AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg	179
Db	2626	AATCTAAAGCCGTCGAA--ATAAGCAAGGGGGGAGTCTTCTGTAGTATGTTAATGGA	2682
Qy	180	LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly---	198
Db	2683	GCAGTAGCTGTTCTGTAAACATTTCAAAGCAAAATCAGCAGCCGATATCTTAGAAGAGAC	2742
Qy	199	LysAsnPheValLeuHisThrGluGlyLeuTyrArg-----	210
Db	2743	AAATCATGGGAATTCAAATCAAATAAGCTTTATTCAAGCAAAAATAAAGCTTTACCCAT	2802
Qy	211	-----LysSerGlyAspTyr-----AlaValProArgTyrArg	221
Db	2803	TCTTTAGCTAGCTGGAACCAAGGGGATTTGACGGGGTCCCATTTATATCTCAACGA	2862
Qy	222	AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg---Ala	240
Db	2863	AAT-----TCAATTGAACCCCAAGTCCCATAAAGATGCA	2895
Qy	241	ValLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr	260
Db	2896	TTAAAGGCGCTCAAAAGT-----TATCATPCGATTAAATCGCCAAACACAGAGATCAA	2946
Qy	261	GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTrpGlnIy	280
Db	2947	-----TCTGCATACITTTGGATGCAAGATGAGTGTCCAAAGCCAGATGATTATAACAGT	3000
Qy	280	sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro-----	292
Db	3001	TGTTTACCTTTCGCCAAACACCTCGGATTTTATCTCCCAAAGAGAAACCGTAAACGTT	3060
Qy	293	-----HisIleLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCy	309
Db	3061	TCAGATTATACGGGGGTACCCGTATCAACCTTAATCCAATGAATATGAAGACCGAGTCT	3120

QY	309	sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpLeuAs	329
DB	3121	TGGTTTTAAAGAGAGGGGTAT-CATTTTTCTGAACCAACATATATTGGTGCGTATTTTGA	3179
QY	329	pLeuArgAsnLysArgTyrgluLeuArgAlaGluTrpLysGlnPropheProglyPheGl	349
DB	3180	ATTCAACAACAAAAAATTGATATCCGT-----GATATGACATTTCCCGCTTAT--	3228
QY	349	uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAl	369
DB	3229	-----TTAAGATCAACAGAAAACCGGATGATAG	3257
QY	369	aValgluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgilecIeuLeuArgHi	387
DB	3258	CAGTGGCTCTTTTATCCAAAGCAAGATTATGTCGCATATCAACGTATTAG-----	3309
QY	387	sGlnProilegIyArg-----le	393
DB	3310	-----GATGCCGCGAGGGTTACTATGCAATGGGCTTTATTTCGATGAACACCATAG	3362
QY	393	uLysGlySerTrpGlyValGlnTyrluLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe	413
DB	3363	AAACACGGCTGTAGTATTGAATATATTTACGAAATTAAGAACAAGCGGCATCATTGA	3422
QY	413	rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis--	428
DB	3423	CAAAGCAGTGTAAAGTCTAATCAACAAAACATCATACTTGACAGTTATATGCAACATAC	3482
QY	429	-----GlnAlaLenTrp-----Ty	429
DB	3483	GCATTGCAGTCTTTATCCTAATCCAAGTAGAATGCCGCCAACACGCGATAAAACCTTA	3542
QY	429	rSerPhePheGlyValGlu-----	435
DB	3543	TTCATACTATCATCTGATAGAATGTTTATAAGAAAACATAATATGTTGCAATTGAA	3602
QY	436	-----GlnAlaLenTrp-----	439
DB	3603	TTTAGAGAAAAAATTCAACAAAATTGGCTTACTCATCAAAATGTCTCAATCTTGTT	3662
QY	440	AspAsnPheThr-----LeuGluGlyValargValGl	451
DB	3663	TGATGACTTTACTTCAGCGCTTCAGCATAAAGATTATTTAACTCGACGTTTACCGCTAC	3722
QY	451	uLysGlnLysAlaSerIleArgTyrglyAspLysAlaLeulleAspArgGluasnTyrrily	471
DB	3723	GGCAACAGAGTATTCA-----GAGAAAGCTAATGAAACAGAGAAATGGTTACAA	3773
QY	471	sGln-----ProLeuProAspLeugly-----	478
DB	3774	AAAAACAACCTTACTTATATCCAAAACCAACAGTAGGTTTTGTAGTACAAGATCATTTGTA	3833
QY	479	-----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs	492
DB	3834	TTATAAGGTAACTCCTTAATTACAGAGACTGTAAGTGC GGTTAATTAAAGGGA AAAA	3893
QY	492	nTrpTyPheThrProGlnHisLysLeuSerLex-----	503
DB	3894	TTATTATTCGACGACGCAATAATATGCAATTAGGAAATACGTTGATTAGGTTTAGG	3953
QY	504	-----ThrAlaSerHisGlnGlu-----	509
DB	3954	TATTTCGTATGACGTATCTCGCAACAAAGCTAATGAATCACTATTAGTGTGGTAAATT	4013
QY	509	-----	509
DB	4014	TAAAAATTTCTCTTGGAACTCTGGTATTGTGTCATNAACCAACCGGAATGGCTTGATCTTTC	4073
QY	510	-----ArgLeuProSerThrGlnGlnLysTyrrAlaHisGlyLy	522
DB	4074	TTATCGCCTTCTACTGGATTTAGAAATCTTAGTTTTCGTGAATGTAT-----GGTTG	4127
QY	522	sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs	542

```
Db 4128 GCGTATGTCGCGCAATAATACGAGGTTATGTAGTAAATTTAAAGCTGAAACATCTCG 4187
Qy 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTy 562
Db 4188 TAACCAAGAGTTTGGTCTCGCTCTAAAGGGGATTTGGTAATATTGAGATCATGTCATT 4247
Qy 562 rAspAsnArgPheGlyAsnTyrIle---TyrAlaGlnThrLeuAsn-----AspG 578
Db 4248 TAGTAATGCTTATCGAAATCTTATCGCCTTTTGCTGAAGAACTTAATAAAATGGAAC 4307
Qy 578 yArgGlyProIlySerIleGluAspAspSerGluMetIlySerValArgTyrAsn---G 597
Db 4308 AAAGCCAAATATCGATATCATATATGTCACAAATGCAAAATATTAGTGGCGTAATAATA 4367
Qy 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
Db 4368 TGGCAATTAGATTTTAATGCTTTATGGAACGATTCCTACGGTTGGTATGCAACATT 4427
Qy 611 ----- 611
Db 4428 TGCTTATAACCGAGTAAAGTTAAAGATCAAAAATCAATCTGCTTTGGCTCGGTAAG 4487
Qy 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTTATTTGATGCCATTCAGCCAGC---CGTTATATCATTTGTTTAGGCTATGA 4544
Qy 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCCAAAGTAATACTTGGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCA 4604
Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTTGCTAGGAACGCT---GCATTGGGTAACAAT-----TCAAGGATGT 4652
Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAAATCAACAAGAAACTTACTCGGCA-----TGGCATATC----- 4689
Qy 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGTATCGGTTATTACATGGCGAATAAA----- 4722
Qy 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 4723 -----AATATTATCTTCGATTAGGATATATAATTTATTCAA 4760
Qy 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLys---AlaAspAsnLeuLeuAsn----- 730
Db 4761 CTATCGCTATGTTACTTTGGGAAGCGGTGCGTCAAAACAGCACAAAGTGGGTCAATCAACA 4820
Qy 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATCTACTCGCTACGCAGCATCA-----GCACGAAACTA 4868
Qy 750 eThrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTAGAAATGAAATTC 4893

RESULT 27
US-08-478-373-4
; Sequence 4, Application US/08478373
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
```

```
;
;
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(160..2121, 2152..4890)
; US-08-478-373-4

Alignment Scores:
Pred. No.: 4,53e-08 Length: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
DB: 2 Gaps: 48

US-09-936-377-2 (1-758) x US-08-478-373-4 (1-5099)

Qy 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 20
Db 2155 ACTAAAAAACCCCTATTTCGCTAAAGTATTATTCTTGTCTTTTAAATTCATGCTATGTA 2214

Qy 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
Db 2215 AAAGCAGAAACTCAAGTATAAAAGTACAAAGAACGCTATATCATCTGAAAGTGGACACT 2274

Qy 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
Db 2275 CAAAGTACAGAAAGATTTCAGAAATTAGAAACTATCTCAGTCTACTGCAGAAAAAATAAGAGAT 2334

Qy 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2335 CGTAAAGATAAATGAAGTA-----ACTGGACTTCGCAAAATTTATCAAAACTAGT 2382

Qy 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
```

2383	Db	GAAAGTATCAGCCGAGAACAAAGTATTAAATATTTCGTGATCTAAACAGCTATGATCAGGC	2444
83	Qy	ILEHISAlaSerGlnTyrGlyGlyGlyAlaSerAlaProVal---ILEargGlyGlnThr	101
2443	Db	ATTTCAAGTTCTAGAACAAAGCCGTGGTGCAGTTCTGGATATTTCTATTCGTGCTATGGAC	2502
102	Qy	GlyArgGlnIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaaspPheSer	121
2503	Db	AGAAATAGAGTT-----	2514
122	Qy	ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGly	141
2515	Db	-----GCTTATTAGTAGAGTTTACCTCAAACGCAATCTTATGTAGTGCAGAACG	2565
142	Qy	ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValaspValAla	159
2566	Db	CCITTAGTTCGTCTCAGATATTTTCGCACATTCGTGCAATTAATGAATTAATGAATGA	2625
160	Qy	AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg	179
2626	Db	AATGTAAGCCGTCGAA---ATAAGCAAGGGGGGAGTCTTCTGAGTATGTAATGGA	2682
180	Qy	LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly---	198
2683	Db	GCACAGTCTGCTTCTGTAAACATTTCAAAGCAAAATCAGACGCCGATATCTTTAGAAGGAGAC	2742
199	Qy	LysAsnPheValLeuHisThrGluGlyLeuTyrArg-----	210
2743	Db	AAATCATGGGGANTTCAACTAATAAATGCTTATTCAAGCAAAATAAAGGCTTTACCCAT	2802
211	Qy	-----LysSerGlyAspTyr-----AlaValProArgTyrArg	221
2803	Db	TCITTAGCTGTAGTCGAAACAAAGGGGATTTGACGGGTGCGCCATTTATCTCAACGA	2862
222	Qy	AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg---Ala	240
2863	Db	AAT-----TCAATTGAAACCCCAAGTCCCATTAAGATGCA	2895
241	Qy	ValLeuGlyTTPArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyr	260
2896	Db	TTAAAGGCGTCAAAAGT-----TATCATGATTATCGCCAAACAGAGGATCAA---	2946
261	Qy	GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTrpGlnLys	280
2947	Db	-----TCTGCATATCTTTGTGATCAAGATGAGTGTCCAAAGCCAGATGATTATAACAGT	3000
280	Qy	sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro-----	292
3001	Db	TGTTTACCTTTCGCCAAACGACCTGCGATTTTATCTCTCCCAAGAGAAACCGTAAGCGTT	3060
293	Qy	-----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys	309
3061	Db	TCAGATTATACGGGGGTAAACCGTATCAAACCTAATCCAATGAATAATGAAGCCAGTCT	3120
309	Qy	sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAs	329
3121	Db	TGGTTTTTAAGAGGGGGGTAT-CATTTTTCTGAACAACATTATATTCGGTGGTATTTTTGA	3179
329	Qy	pLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheG	349
3180	Db	ATTCCACACAAAAAATTGATATCCGT-----GATATGACATTTTCCCGCTTAT--	3228
349	Qy	uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAl	369
3229	Db	-----TTAAGATCAACAGAAAACGGGATGATAG	3257
369	Qy	aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi	387
3258	Db	CAGTGGCTCTTTTATCCAAAGCAAGATTATGGTGCATATCAACGTATTCAG-----	3309
387	Qy	sGlnProIleGlyArg-----Le	393
3310	Db	-----GATGCGCCGCGCTTAACATATCGAAGTGGGCTTTATTTTCGATGAACACCATAG	3362

[illegible]


```
QY 611 ----- 611
Db 4428 TGCTTATAACCGAGTAAAGTAAAGTCAAAAATCAATCTGGTTGGCCTCCGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTATTGTGATGCAATCAGCCAGC---CGTTATATCATGTTTAGCTATGA 4544
QY 624 P-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCCAGTAATACTTTGGGAATTAATACAAATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTCGTAGAAAACGT---GCATTGGGTAAACAT-----TCAAGGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAAATCAACAAGAAAACCTTACTCGGCA-----TGGCATATC----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGTATCGGGTTATTACATGCGCAATAAA----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
Db 4723 -----AATATATGCTTCGATTAGGGATATATATTTATTCAA 4760
QY 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLys---AlaAspAsnLeuLeuAsn----- 730
Db 4761 CTATCGCTATGTTACTTTGGGAAGCGGTGCGTCAAAACAGCACAGGTGCGGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATCTCCTACGCGAGCATCA-----GGACGAAACTA 4868
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTAGAAATGAATTC 4893

RESULT 28
US-08-474-671-4
; Sequence 4, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
```

```
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(160..2121, 2152..4890)
; US-08-474-671-4

Alignment Scores:
Pred. No.: 4.53e-08 Length: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
DB: 3 Gaps: 48

US-09-936-377-2 (1-758) x US-08-474-671-4 (1-5099)

QY 5 ThrLeuLysProLe-----ValLeuSerIleLeuLeuLeuAsnThrProLeu 20
Db 2155 ACTAAAAAACCCCTATTTTCGCTAAAGTATTATTCTTCTTTAAATTCATGCTATGTA 2214
QY 21 LeuAlaGlnAlaHis-----GluThrGlnSerValGly----- 32
Db 2215 AAAGCAGAAACTCAAGTATAAAGATACAAAGAACGTATATCATCTGAAGTGGACACT 2274
QY 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
Db 2275 CAAAGTACAGAGATTCCAGATTAGAACTATCTCAGTCTACTGCAGAAAAAATAGAGAT 2334
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2335 CGTAAAGATAAATGAAGTA-----ACTGGACTTGGCAAAATTTATCAAAACTAGT 2382
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2383 GAAAGTATCAGCCGAGCAAGCAAGTATTAAATTCGTGATCTAAACACGCTATGATCCAGGC 2442
QY 83 IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr 101
Db 2443 ATTTCAGTTGTAGAACAAAGCCCGTGTGCAAGTCTCGATATTCTATTCTGGTATGGAC 2502
QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
Db 2503 AGAAATAGAGTT----- 2514
QY 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGly 141
Db 2515 -----GCITTATTAGTAGTGTTTTACTCTCAACCGCAATCTTATTATGTAGTCAAGC 2565
QY 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
Db 2566 CCITTAGTTGCTGTTTCAGGATATTCTGGCAGTGGTGCATTAATGAAATTGATATGAA 2625
QY 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuArg 179
```

Db 2626 AATGTAAGGCGCGTGAA---ATAAGCAAGGGGGAGTCTCTCGAGTATGTAATGGA 2682
QY 180 LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly--- 198
Db 2683 GCATAGCTGTGTTCTGTAAACATTTCAAGCAAAATCAGCAGCGGATATCTTAGAAGGAGAC 2742
QY 199 LysAsnPheValLeuHisThrGluGlyLeuTyrArg----- 210
Db 2743 AAATCATGGGAATTCAAACTAAATAATGCTTATCAAGCAAAATAAAGGCTTTTACCAT 2802
QY 211 -----LysSerGlyAspTyr-----AlaValProArgTyrArg 221
Db 2803 TCTTTAGCTGTAGCTGGAAACCAAGGGGATTTGACGGGTGCGCATTTATATCAACGA 2862
QY 222 AsnLeuLysArgLeuProAspSerProArgPheAlaAsnGlyGlnHisArg---Ala 240
Db 2863 AAT-----TCAATTGAAACCAAGTCCATAAAGATGCA 2895
QY 241 ValLeuGlyTyrArgPheTyrArgThrTyrSerAspArgAspGlnTyr 260
Db 2896 TTAAGGCGGTACAAGT-----TATCATCGATTAAATCGCCAAACCCAGAGGATCAA--- 2946
QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTyrGlnLys 280
Db 2947 -----TCTGCATATTTGTGATGCAAGATGAGTGTCCAAAGCCAGATGATATACAGT 3000
QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292
Db 3001 TGTTTACCTTTGCGCAAAACGCTCGATTATATCTCCCAAGAGAAACCGTAAGCGTT 3060
QY 293 -----HisLeuLeuThrGluLysValAspTyrAspAsnProGlyLeuSerCy 309
Db 3061 TCAGATTATACGGGGCTTAACCGTATCAACCTTAATCCAAATGAATATGAAGCCAGTCT 3120
QY 309 sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAs 329
Db 3121 TGGTTTTTAAGAGGAGGTAT-CATTTTCTGCAACAACATTATATGTTGGTATTTTGA 3179
QY 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheG 349
Db 3180 ATTCAACAACAAAAATTTGATATCCGT-----GATATGACATTTCCGCTTAT-- 3228
QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAl 369
Db 3229 -----TTAAGATCAACAGAAAACCGGATGATAG 3257
QY 369 aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi 387
Db 3258 CAGTGGCTCTTTTATCCAAAGCAAGATTATGGTGCATATCAACGTATTGAG----- 3309
QY 387 sGlnProIleGlyArg----- 393
Db 3310 -----GATGGCGAGCGGTTAACTATGCAAGTGGGCTTTATTTGCGATGAACACCATAG 3362
QY 393 uLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
Db 3363 AAAACAGCGTGTAGTATGTAATATATTTACGAAATAAGAACAAAGCGGCATCATGA 3422
QY 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis-- 428
Db 3423 CAAAGCAGTGTAAAGTGTATCAACAAACATCATCTTGCACAGTTATATGCAACATAC 3482
QY 429 ----- 429
Db 3483 GCATTGCGAGTCTTATCTCTAATCCAGTAAGATTGCGGCCCAACACGATGAACCTTA 3542
QY 429 rSerPhePheGlyValGlu----- 435
Db 3543 TTCATCTACTCATCTGTAGATAAATGTTTATAAGAAAAACATAATATGTTGCAATTGAA 3602
QY 436 -----GlnAlaAsnTrp----- 439

Db 3603 TTTAGAGAAAAAATTCACAAAAATTTGGCTTACTCATCAAAATTTCTTCAATCTTGGTTT 3662
QY 440 -AspAsnPheThr-----LeuGluGlyGlyValArgValG 451
Db 3663 TGATGACTTTACTTTCAGCGCTTTCAGCATAAAGATTATTAACTCGACGTTTACCCTAC 3722
QY 451 uLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLy 471
Db 3723 GCGAAAGAGATTTTCA-----GAGAAAGCTAATGAAAAACAGAAAGAAATGGTTACAA 3773
QY 471 sGln-----ProLeuProAspLeuGly----- 478
Db 3774 AAAACAACCTTACTTATATACCCAAACACAGTAGCTTTTGTAGTACAGATCATTTGTA 3833
QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAAGTAACTCTCTTAATTACAGAGACTGTAAAGTGGCGTTAAATTAAGGGNAAAA 3893
QY 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu----- 503
Db 3894 TTATTATTTCGACGACGCAATAATATGTCATTAGGAAATACGTTGATTAGGTTTAGG 3953
QY 504 -----ThrAlaSerHisGlnGlu----- 509
Db 3954 TATTCGGTATGACGTATCTCGCACAAAAGCTAATGAATCAACTATTAGTGTGGTAAAT 4013
QY 509 ----- 509
Db 4014 TAAAAATTTCTTCTTGAATACTGGTATTGTTCATAAAACCAACGGAATGCTTGAATCTTC 4073
QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
Db 4074 TTATCGCCTTTCTACTCGAATTAGAAATCTTAGTTTTCGCTGAATGTAT-----GGTTG 4127
QY 522 sHisValAlaThrAsnThrPheGluValGlyLysAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 CGCGTATGTCGCAATAATAGCAGGTTTGTAGTAAATTTAAGCCCTGAACATCTCG 4187
QY 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyr 562
Db 4188 TAACCAAGAGTTTGTCTCGCTCTAAAGGGGATTTTGGTAATATTGAGATCAGTCATTT 4247
QY 562 rArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsn-----AspG 578
Db 4248 TAGTAATCTTATCGAAATCTTATCGCCTTTGCTGAGAACCTTAATAAAAAATGGAAC 4307
QY 578 YArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn---G 597
Db 4308 AAAGGCCAATTATGGATATCATATGACAAAATGCAAAATTAGTTGGGTAAATATAAC 4367
QY 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
Db 4368 TCGCAATTAGATTTTAAATGTTTATGAAACGTTATTCCTACGGTTGTTATGCAACATT 4427
QY 611 ----- 611
Db 4428 TGCTTATAACCGAGTAAAGTTAAAGATCAAAAAATCAATGCTGTTGGCTCCGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTATTATGATGCCATTTCAGCCGAGC-----CGTTATATCATTTGTTAGGCTATGA 4544
QY 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCCAAGTAATACTTGGGGAATTAATAAATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATGCTAGGAAAAACGT---GCATTGGGTAAACAAT-----TCAAGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAAGAAACTTACTCGGCA-----TGGCATATC----- 4689


```

Db      2896  TTTAAAGGGGTACAGT-----TATCATGATTAATCGCCAAACACGAGATCAA--- 2946
QY      261  GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTyrGlnly 280
Db      2947  -----TCTGCATACTTGTGTGATGCAAGATGAGTCCAAAGCCAGATGATTATAACAGT 3000
QY      280  sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292
Db      3001  TGTTTACCTTCGCCAACGACCTCGGATTTTATCTCTCCAAAGAGAAACCGTAAGCGTT 3060
QY      293  -----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCy 309
Db      3061  TCAGATTATACGGGGCTAACCGTATCAACCTAATCAATCAATCAATCAATCAATCAATCAAT 3120
QY      309  sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAs 329
Db      3121  TGTGTTTTTAAAGGGGGGTAT-CATTTTCTGACACACATATATATGTTGGTATTTTGA 3179
QY      329  pLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrPlysGlnProPheProGlyPheG 349
Db      3180  ATTCACACAAACAAAATTTGATATCGT-----GATATGACATTTCCCGCTTAT-- 3228
QY      349  uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAl 369
Db      3229  -----TTAAGATCAACAGAAACACGGGATGATAG 3257
QY      369  aValGluAsnPheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi 387
Db      3258  CAGTGGCTCTTTTATCCAAAGCAAGATTATGTCGATATCAACGTTATTGAG----- 3309
QY      387  sGlnProIleGlyArg-----Le 393
Db      3310  -----GATGGCGGAGGGTTAACTATGCAAGTGGGCTTTATTTTCGATGAACACCATAG 3362
QY      393  uLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
Db      3363  AAAACAGCGTGATGATTTGAATATATTTACGAAATAGAACAAAGCGGGCATCATGCA 3422
QY      413  rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis-- 428
Db      3423  CAAAGCAGTGTAAAGTGCTAAATCAACAAACATCATCTACTTGACAGTATATATGCAACATAC 3482
QY      429  -----GlnAlaAsnTrp-----Ty 429
Db      3483  GCATTCGAGCTTTATCTTAATCCAAAGTAAGATTCGCCGCCCAACACGTGTATAACCTTTA 3542
QY      429  rSerPhePheGlyValGlu----- 435
Db      3543  TTCATCTATCATCTCTGATAGATAAATGTTTAAAGAAACAAATATATGTTGCAATTGAA 3602
QY      436  -----GlnAlaAsnTrp----- 439
Db      3603  TTTAGAGAAAAAATTCACAAATTTGGGTTACTCATCAAAATGCTTCAATCTCTTGGTTT 3662
QY      440  -AspAsnPheThr-----LeuGluGlyValArgValG 451
Db      3663  TGATGACTTTACTTCAGCGCTTCAGCATAAAGATTATTTAACTCGACGTTTACCGCTAC 3722
QY      451  uLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLy 471
Db      3723  GGCAAAAGAGTATTTC-----GAGAAAGCTAATGAACAAAGAAATGTTTACAA 3773
QY      471  sGln-----ProLeuProAspLeuGly----- 478
Db      3774  AAAACACCTTACTTATACCCAAACCAACAGTAGTGTGTTTGTAGTACAAGATCATTTGCA 3833
QY      479  -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db      3834  TTATAAAGGTAACTCTCTAATTACAGAGACTGTAAGTGGGTTAATTAAGGGGAAAAA 3893
QY      492  nTyrTyrPheThrProGlnHisLysLeuSerLeu----- 503

```

```

Db      3894  TTATTATTTCGCAGCACCAATAATATGTCATTAGGGAATACGTTTATTAGTTTAGG 3953
QY      504  -----ThrAlaSerHisGlnGlu----- 509
Db      3954  TATTTCGGTATGACGTATCTCGCAAAAGCAATGATGAATCAACTATTAGTGTGGTAAATT 4013
QY      509  ----- 509
Db      4014  TAAAAATTTCTCTTGGGAATACTGCTATTGTCTATAAAACCAACGAATGGCTTGATCTTC 4073
QY      510  -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
Db      4074  TTAFCGCGCTTTCTACTCGGATTTAGAAATCCTAGTTTGTGGAATGAT-----GGTTG 4127
QY      522  sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
Db      4128  CGCGTATGCTGGCATAATATAGCGAGGTTTATGAGGTAAATTTAAGCCTGAACATCTCG 4187
QY      542  nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTy 562
Db      4188  TAACCAAGAGTTTGGTCTCGCTCTAAAGGGGATTTTGGTAATATTGAGATCAGTCAATT 4247
QY      562  rArgAsnArgPheGlyAsnTyrIle---TyrAlaGlnThrLeuAsn-----AspGl 578
Db      4248  TAGTAATGCTTATCGAAATCTTATCGCCTTGTCTGAAGAATTTAATAAATGGAATCGG 4307
QY      578  yArgGlyProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsn---Gl 597
Db      4308  AAAGCCCAATTTATGGATATCATATGCAAAATGCAAAATAGTTAGTGGCGTAAATATAAC 4367
QY      597  nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
Db      4368  TGCAGCAATTAGATTATTAATGGTTTATGGAACAGTATTCCTACGGTGGTATGCAACATT 4427
QY      611  ----- 611
Db      4428  TGCATTATAACCGAGTAAAGTTAAAGATCAAAATCAANTCACTGCTGTTTGGCTCCGTAAG 4487
QY      612  -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db      4488  CAGTTATTATTATTTGATGCCATTCAGCCAGC---CGTTATATCATCTGTTTAGGCTATGA 4544
QY      624  p-----TyrValArgGlyArgLeuLysAsnLe 633
Db      4545  TCATCCAAAGTAATACTTGGGGAATTAATACATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY      633  uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db      4605  AAATGCAATTGTAGGAAACCGT---GCATTGGGTAAACAAT-----TCAAGGGATGT 4652
QY      653  pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db      4653  AAAATCAACAAAGAAACTTACTCGGCA-----TGGCATATC----- 4699
QY      673  rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db      4690  -----TTAGATGATCGGTTATTACATGCGGATAAA----- 4722
QY      693  gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db      4723  -----AATATTATGCTTCGATTAGGGATATATATTTATTATTCAA 4760
QY      713  nThrArgTyrGlyGluTyrPheAsnTyrTyrValLys---AlaAspAsnLeuLeuAsn--- 730
Db      4761  CTATCGCTATGTTTACTTGGGAAGCGGTGCGTCAACAGCACAGAGTGGGTCATATCAACA 4820
QY      731  -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db      4821  TCAAAATGTTGGTAGCTATCTCGCTACGCGAGCATCA-----GGACCAAACTA 4868
QY      750  eThrGlyGlyValAsnValLysPhe 758
Db      4869  TACCTTAACATTAGAAATGAAATTC 4893

```

RESULT 30
 US-08-897-438-4
 ; Sequence 4, Application US/08897438
 ; Patent No. 6262016
 ; GENERAL INFORMATION:
 ; APPLICANT: Loomore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 160
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/897,438
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/483,577
 ; FILING DATE: 07-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/337,483
 ; FILING DATE: 08-NOV-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/175,116
 ; FILING DATE: 29-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/148,968
 ; FILING DATE: 08-NOV-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-720
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5099 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: join(160..2121, 2152..4890)
 US-08-897-438-4

Alignment Scores:
 Pred. No.: 4,538-08 Length: 5099
 Score: 180.50 Matches: 197
 Percent Similarity: 33.30% Conservative: 139
 Best Local Similarity: 19.52% Mismatches: 323
 Query Match: 4.47% Indels: 351
 DB: 3 Gaps: 48

US-09-936-377-2 (1-758) x US-08-897-438-4 (1-5099)

QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20

Db	2155	ACTAAAAAACCCCTATTTTCGCCTAAGTATTATTTCTTGCTTTTAATTTTCATGCTAGTA	2214
QY	21	LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly-----	32
Db	2215	AAAGCAGAAACTCAAAAGTATAAAGACTCAAAAGAGCTATATCATCTCGAAGTGGAACACT	2274
QY	33	-----LeuGluThrValThrValValGly--LysSerArgPro	44
Db	2275	CRAAGTACAGAGATTGAGAACTATCTCAGTCAGTCAGCAGAGAAAATAAGAGAT	2334
QY	45	ArgAlaThrSerGlyLeuHisThrSerThrAlaSerAspLysIleIle	62
Db	2335	CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATATCAAACTAGT	2382
QY	63	GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly	82
Db	2383	GAAAGTATCAGCCGAGAACAAAGTATTAAATATTCGTGATCTAACACGCTATGATCCAGC	2442
QY	83	IleHisAlaSerGlnTyrglyGlyAlaSerAlaProVal--IleArgGlyGlnThr	101
Db	2443	ATTTTCAGTTGTAGAACAGGCGGTGCAAGTTCTGGATATCTTCTGCTGTATGGAC	2502
QY	102	GlyArgArgIleLysValIleAsnHisGlyGluThrGlyAspMetAlaAspPheSer	121
Db	2503	AGAAATAGAGTT-----	2514
QY	122	ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValIleLeuArgGly	141
Db	2515	-----GCCTTATTAGTAGATGTTTACCTCAACGCAATCTTATGTCGCAAGC	2565
QY	142	ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla	159
Db	2566	CCTTTAGTTGCTCGTTCAGGATATTTCTGCACCTGGTGAATTAATGAATATGAATGA	2625
QY	160	AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuLeuArg	179
Db	2626	AATGTAAGGCGCTCGAA---ATAAGCAAGGGGGGAGTCTCTCTGAGTATGTAATGA	2682
QY	180	LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly	198
Db	2683	GCACCTAGCTGGTTCGTGTAAACATTTCAAGCAAAATCAGCAGCCGATATCTTAGAAGGAG	2742
QY	199	LysAsnPheValLeuHisThrGluGlyLeuTyrg--	210
Db	2743	AAATCATGGGAATTCAAATGCTTATTCAAGCAAAATAAAGGCTTTACCCAT	2802
QY	211	-----LysSerGlyAspTyr-----AlaValProArgTyrg	221
Db	2803	TCTTTAGCTGTAGCTGGAACAAAGGGGATTTGACGGGGTCCCATTTATCTCAACGA	2862
QY	222	AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg--Ala	240
Db	2863	AAT-----TCAAATGAACCCCAAGTCCCAAGATGA	2895
QY	241	ValLeuGlyTrpArgLysArgPheTyrgArgThrTyrgSerAspArgAspGlnTyrg	260
Db	2896	TTAAAGGCGTACAAAGT-----TATCATGATTAATCGCCCAACACGAGATCAA	2946
QY	261	GlyLeuProAlaHisSerHisGluTyrgAspAspCys-HisAlaAspIleIleTrpGlnly	280
Db	2947	-----TCTGCATACTTTGTGATGCAAGATGTCGCAAGCCAGATGATTATAACAGT	3000
QY	280	SerLeuIleAsnLysArgTyrgTyrgLeuGlnLeuTyrgPro-----	292
Db	3001	TGTTTACCTTTGCGCAACAGACCTGCGATTTTATCTCTCCCAAGAGAAACCGTAAGCGTT	3060
QY	293	-----HisLeuLeuThrGluLysValAspTyrgAspAsnProGlyLeuSerCy	309
Db	3061	TCAGATTATACGGGGCTAACCGTATCAACCTAATCAATGAATATGAAGCCAGTCT	3120
QY	309	sglyPheHisAspAspAlaHisAlaHisAsnGlyLysProTrpIleAs	329

Db 3121 TGGTITTTTAAAGAGGGGTAT-CATTTTCTGAACAACATTATATGCTGGTATTTTGA 3179
 Qy 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheCl 349
 Db 3180 ATTACACACAAACAAATTTGATATCGT-----GATATGACATTTCCCGCTTAT-- 3228
 Qy 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAl 369
 Db 3229 -----TTAAGATCAACAGAAAAACGGGATGATAG 3257
 Qy 369 aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHis 387
 Db 3258 CAGTGGCTCTTTTCCAAAGCAACATATTGCTGCATATCAACGTTATTGAG----- 3309
 Qy 387 sGlnProIleGlyArg-----Le 393
 Db 3310 -----GATGGCCGAGGCGTTAACTATGCAAGTGGGCTTTATTTTCGATGAACACCATAG 3362
 Qy 393 uLysGlySerTrpGlyValGlnTrpLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
 Db 3363 AAAACAGCGGTAGGTATTGAATATATTACGAAAAATAAGAACAAAGCGGCATCATTTGA 3422
 Qy 413 rGluAlaVal-----LysGlnProMetLeuLeuAspAsnLysValGlnHis-- 428
 Db 3423 CAAAGCAGTGTAAAGTCTTAATCAACAACATCATACTTTGACAGTTATATGCAACATAC 3482
 Qy 429 -----Ty 429
 Db 3483 GCATTGCGCTTTTATCTTAATCCAAAGTAAAGATTGCGGCCCAACACGCGTATAAACCTTA 3542
 Qy 429 rSerPhePheGlyValGlu----- 435
 Db 3543 TTCATACTATCATCTCTGATAGAAATGTTTATAAGAAAAACATAATATGTTGCAATTGAA 3602
 Qy 436 -----GlnAlaAsnTrp----- 439
 Db 3603 TTTAGAGAAAAAATTCAACAAATTTGGCTTACTCATCAATTTGCTTCAATCTTGTTT 3662
 Qy 440 -AspAsnPheThr-----LeuGluGlyValArgValG 451
 Db 3663 TGATGACTTTACTTCAGCGCTTCAGCATAAAGATTATTTAACTCGACGTGTTACCGCTAC 3722
 Qy 451 uLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTrpLys 471
 Db 3723 GGCNAAGAGTATTTCA-----GAGAAAGCTTAATGAACAAAGAAATGTTTACAA 3773
 Qy 471 sGln-----ProLeuProAspLeuGly----- 478
 Db 3774 AAAACAACCTTACTTATACCCMAAACCAACAGTAGGTTTTGTAGTACAAAGATCATTTGTA 3833
 Qy 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
 Db 3834 TTATAAAGGTAACTCTCTAATTACAGAGACTGTAAAGTCGCGTTAATTAAAGGGAAAAA 3893
 Qy 492 nTriPyrPheThrProGlnHisLysLeuSerLeu----- 503
 Db 3894 TTATTATTTCGACGACGCAATAATATGCGATTAGGGAATACGTTGATTAGTTTAGG 3953
 Qy 504 -----ThrAlaSerHisGlnGlu----- 509
 Db 3954 TATTGCGTATGACGTATCTCGCACAAAGCTAATGATCAACTATTAGTGTGTTAAAT 4013
 Qy 509 ----- 509
 Db 4014 TAAAAATTTCTTGGATACTGTTATGTCATAAAAACCAACGAATGGCTTATCTTTC 4073
 Qy 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLys 522
 Db 4074 TTATCGCTTTCTACTCGAATTAGAAATCTAGTTTTCGCTGAAATGAT-----GTTG 4127
 Qy 522 sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
 Db 4128 GCGGTATGGTGGCAATAATAGCGAGGGTTTATGTAGGTAAATTTAAGCCCTGAAACATCTCG 4187

Qy 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTy 562
 Db 4188 TAAACAGAGTTTGGCTTCGCTCTAAAGGGGATTTTGGTAATATTAGATCACTCATTT 4247
 Qy 562 rArgAsnArgPheGlyAsnTrpIle-----TyrAlaGlnThrLeuAsn-----AspG 578
 Db 4248 TAGTAATGCTTATCGAAATCTTATCGGCTTTGCTGAAGAACTTAATAAAATGAACCTGG 4307
 Qy 578 yArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn---G 597
 Db 4308 AAAAGCCCAATTATGATATCATATCAATAAGTCAAAATAGTTCGGCGTAAATATAAC 4367
 Qy 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe----- 611
 Db 4368 TGGCAATTAGATTTTAATGCTTTATGMAACGTTATCCCTACGCTGGTGGTATGCAACATT 4427
 Qy 611 ----- 611
 Db 4428 TGCTTATAACGAGTAAAGATTAAAGATCAAAAATCAATGCTGTTGGCTCCGTAAG 4487
 Qy 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
 Db 4488 CAGTTATTTATTTGATGCCATTCAGCCGACC-----CGTTATATCATTTGTTTAGCTATGA 4544
 Qy 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
 Db 4545 TCATCCAAAGTAATACTTGGGGAATTAAATACATGTTTACTCAATCAAAAGCAAAATCTCA 4604
 Qy 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
 Db 4605 AAATGAATTTGCTAGAAACCGT---GCATTGGGTAAACAAAT-----TCAAGGGATGT 4652
 Qy 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
 Db 4653 AAAATCAACAAGAAACTTACTCGGCA-----TGGCATATC----- 4689
 Qy 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
 Db 4690 -----TTAGATGTATCGGGTTATTACATGCGCAATAAA----- 4722
 Qy 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
 Db 4723 -----AATATTATGCTTCGATTAGGATATATTAATTTATTTCA 4760
 Qy 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLys---AlaAspAsnLeuLeuAsn----- 730
 Db 4761 CTATCGCTTATGTTACTTGGGAAGCGGTGCTCAAAACAGACACAAAGTGGGTCAATCAACA 4820
 Qy 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
 Db 4821 TCAAAATGTTGTAGTATATCTCGCTACGACGATCA-----GGACGAAACTA 4869
 Qy 750 ethrGlyGlyValAsnValLysPhe 758
 Db 4869 TACCTTAACATTAGAAATGAATTC 4893

RESULT 31

US-08-637-654-4
 ; Sequence 4, Application US/08637654
 ; Patent No. 6358727
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M
 ; APPLICANT: Harkness, Robin E
 ; APPLICANT: Schryvers, Anthony B
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Mardin, Andrew D
 ; APPLICANT: Klein, Michel H
 ; TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,654
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/CA94/00616
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-595
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5099 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: join(160..2121, 2152..4890)

US-08-637-654-4

Alignment Scores:
Pred. No.: 5099
Score: 180.50 Matches: 197
Percent Similarity: 33.30% Conservative: 139
Best Local Similarity: 19.52% Mismatches: 323
Query Match: 4.47% Indels: 351
Gaps: 48

US-09-936-377-2 (1-758) x US-08-637-654-4 (1-5099)

QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db 2155 ACTAAAAACCTATTTTCGCTAAGTATTATTCTTCTTTTAAATTCATGCTATGTA 2214
QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
Db 2215 AAGCAGAACTCAAAGTATTAAGATACAAAGAGAGCTATATCATCTGAAAGTGGACACT 2274
QY 33 -----LeuGluThrValThrValValGly-----LysSerArgPro 44
Db 2275 CAAAGTACAGAGATTCAGAACTATCTCAGTCACTGCAGAAATAAAGAGAT 2334
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2335 CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATATCAAAACTAGT 2382
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2383 GAAAGTATCAGCGAGACAGTATTAATATTCGTGATCTTAACACGCTATGATCCAGGC 2442
QY 83 IleHisAlaSerGlnTyGlyGlyAlaSerAlaProVal-----IleArgGlyGlnThr 101
Db 2443 ATTTGAGTTGTAGTAAAGCGCGTGCAGAGTTCTCGGATATCTTATTCGTGTATGGAC 2502
QY 102 GlyArgArgIleLysValLeuAsnHisGlyGlyGluThrGlyAspMetAlaAspPheSer 121
Db 2503 AGAAATAGAGTT----- 2514

QY 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGly 141
Db 2515 -----GCTTTATTAGTAGATGTTTACTCTAAACGCAATCTTATGATGCTAAAGC 2565
QY 142 ProValThrLeuLeu-----TyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
Db 2566 CCTTTAGTTGCTGCTCAGGATATTCGCACTGGTCAATTAATGAATGAATGA 2625
QY 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179
Db 2626 AATGTAAGCGCGTGCAGAA-----ATAAGCAAGGGGGGAGTTCTTCTGAGTATGGTAATGGA 2682
QY 180 LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly 198
Db 2683 GCCTAGTCTGTTCTGTAACTTCAAGCAAAATCAGCAGCGGATATCTTAGAAGGAGAC 2742
QY 199 LysAsnPheValLeuHisThrGluGlyLeuTyrArg----- 210
Db 2743 AATCATGGGAATTCAAAATAAATGCTTATTCAAGCAAAAATAAAGGCTTTACCCAT 2802
QY 211 -----LysSerGlyAspTyr-----AlaValProArgTyrArg 221
Db 2803 TCTTTAGCTGTAGCTGGAACCAAGGGGATTTGAGGGGTCGCCATTTATCTCAACGA 2862
QY 222 AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg---Ala 240
Db 2863 AAT-----TCAATTGAAACCCCAAGTCCATAAAGATGCA 2895
QY 241 ValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyr 260
Db 2896 TTAAGAGCGGTACAAAGT-----TATCATCTAATTAATCGCCAAACCGAGAGATCAA 2946
QY 261 GlyLeuProAlaHisSerHisGluTyrAspAspCys-HisAlaAspIleIleTrrpGlnly 280
Db 2947 -----TCGTCATCTTTGTGATGCAAGTGTGTCCTCAAGCCAGATGATTATAACAGT 3000
QY 280 sSerLeuIleAsnLysArgTyrLeuGlnLeuTyrPro----- 292
Db 3001 TGTTTTACCTTTTCGCCAAACGACCTGCGATTTTATCTCTCCAAAGAGAAACCGTAAGCGTT 3060
QY 293 -----HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCy 309
Db 3061 TCAGATTATACGGGGCTAACCGTATCAACCTAATCCATGAATATGAAGACAGCTCT 3120
QY 309 sGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrrpIleAs 329
Db 3121 TGGTTTTTAAGAGGAGGTAT-CATTTTCTGAAACAACATTTATTTGGTGGTATTTTTGA 3179
QY 329 pLeuArgAsnLysArgTyrGluLeuArgAlaGluTrrpLysGlnProPheProGlyPheG 349
Db 3180 ATTCACACAAACAAATTTGATATCCGT-----GATATGACATTTCCCGCTTAT-- 3228
QY 349 uAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAl 369
Db 3229 -----TTAAGATCAACAGAAACAAACCGGATGATAG 3257
QY 369 aValGluAsnPhePheAsnAsnGlnThrGlnAsnAla-----ArgIleGluLeuArgHi 387
Db 3258 CAGTGCCTCTTTTATCCAAAGCAAGATTTGCTGCATATCAACCGTATTGAG----- 3309
QY 387 sGlnProIleGlyArg-----Leu 393
Db 3310 -----GATGCCGAGCGTTAACTATCAAGTGGCTTTTATTTCGATGAACACCATAG 3362
QY 393 uLysGlySerTrrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSe 413
Db 3363 AAAACAGCGGTAGGTATTGAATATATTACCAAAATAAAGAAACGCGGCATCATTTGA 3422
QY 413 rGluAlaVal-----LysGlnProMetLeuLeuAsnLysValGlnHis-- 428
Db 3423 CAAGCAGTGTAAAGTCTAAATCAACAAACATCATCTTACAGTTATATGCAACATAC 3482

QY 429 -----TV 429
Db 3483 GCATTGCACTTTATCTATCTATCCAGTAAGATTGGCGCCCAACACAGTGATATAACCTTA 3542
QY 429 rSerPhePheGlyValGlu 435
Db 3543 TTTCACTACTTCTGATAGATAAATGTTTATAAGAAAAACATAATATGTGCAATTGAA 3602
QY 436 -----GlnAlaAsnTrp 439
Db 3603 TTTAGAGAAAAAATTTCAACAAATTTGGCTTACTCATCAAAATCTCTCAATCTTGGTTT 3662
QY 440 -AspAsnPheThr 451
Db 3663 TGATGACTTTACTTCAGCGCTTCAGCATAAAGATTATTAACTCGAGCTGTACCGCTAC 3722
QY 451 uLysGlnLysAlaSerIleAArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyr 471
Db 3723 GGCAGAGAGATTCTCA -----GAGAAAGCTTAATGAACAAAGAAATGGTTACAA 3773
QY 471 sGln 478
Db 3774 AAACCAACCTTACTTATATACCAACCAACAGTAGGTTTGTAGTACAGATCATTCTGA 3833
QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAAGGTAACTCTCTCTATTAATACAGAGACTGTAAAGTCGGTTAATTAAAGGGAATA 3893
QY 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu 503
Db 3894 TTATTATTTTCGACGACCAATATATGCGATTAGGGAATACGTTGATTAGGTTAGG 3953
QY 504 -----ThrAlaSerHisGlnGlu 509
Db 3954 TATTCGGTATGACGTATCTGCACAAAGCTAATGAATCAACTATTAGTTGTTGTAAT 4013
QY 509 ----- 509
Db 4014 TAAATAATTTCTTGGAACTACTGGTATTGTTCATAAACCAACGAAGTGGTTGATCTTTC 4073
QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
Db 4074 TTATCGCTTTCTACTGGATTTAGAAATCTAGTTTCTGGAATGTAT -----GGTTG 4127
QY 522 shiValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 GCGGTATGGTGCATAATATAGCGAGGTTTATGTAGTAAATTTAGCCCTGAAACATCTCG 4187
QY 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTy 562
Db 4188 TAAACAAGAGTTTGGTCTCGCTCTAAAGAGGAGTTTGGTAATATTGAGATCAGTCAATT 4247
QY 562 rArgAsnArgPheGlyAsnTyrIle -----TyrAlaGlnThrLeuAsn -----AspG 578
Db 4248 TAGTAATGCTTATCGAAATCTTATCGCTTGTCTGAGAACTTAATAAAATGAACTGG 4307
QY 578 yArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn -----G 597
Db 4308 AAAGCCATTATGATATCATATGACCAAAATGCAAAATAGTTGGCGTAATAATAC 4367
QY 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe 611
Db 4368 TCGCAATTTAGATTTTAAATGTTTATGGAACAGTATTCCTACGGTTGGTATGCAACATT 4427
QY 611 ----- 611
Db 4428 TGCATTATACCGAGTAAAGTTAAAGATCAAAAAATCAATGCTGTTGGCTCGCTGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTTTATGTCGCAATTCAGCCAGC -----CGTTATATCATTTGTTAGGCTAGA 4544
QY 624 p -----TyrValArgGlyArgLeuLysAsnLe 633

Db 4545 TCATCAAGTAATACTTGGGAATTAAATACATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTTGCTAGGAAACCGT -----GCATTGGGTAAACAAT -----TCAAGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAAAGAACTTACTCGGCA -----TGGCATATC ----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGTATCGGGTTATTATACATCGCGCAATAAA ----- 4722
QY 693 gTyrGluThrArgThrProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAs 713
Db 4723 -----AATATTATGTTTCGATTAGGATATATAATTTATTCA 4760
QY 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLys -----AlaAspAsnLeuLeuAsn ----- 730
Db 4761 CTATCGCTATGTACTTGGGAACGGTGCCTCAAAACAGCACAGGTGCGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATCTACTCGCTACCGCATCA -----GGAGCAAACTA 4868
QY 750 ethrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTAGAAATGAAATTC 4893

RESULT 32
US-08-649-518-4
; Sequence 4, Application US/08649518
; Patent No. 6361779
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Van-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,518
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968

199	QY	lysAsn	Phe	Val	Leu	His	Thr	Gl	u	G	ly	Leu	Tyr	Arg	-----	210
2743	DB	AAATCAT	GGGAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	CAAAAT	2802
211	QY	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	221
2803	DB	TC	TTT	AGCT	GTAG	TCG	AAAA	CAAG	GGG	GAT	TTG	ACG	GGG	TCG	CCAT	2862
221	QY	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	221
222	QY	Asn	Leu	Gly	Arg	Leu	Pro	Asp	Ser	Pro	Arg	Arg	Phe	Ala	Asn	2862
2863	DB	AAAT	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2863
241	QY	Val	Leu	Gly	Trp	Arg	Gly	Ser	Arg	Phe	Tyr	Ser	Asp	Arg	Arg	2863
2896	DB	TTAA	AGGG	CGT	AC	AAAG	-----	-----	-----	-----	-----	-----	-----	-----	-----	2896
261	QY	Gly	Leu	Pro	Ala	His	Ser	His	Glu	Tyr	Asp	Asp	Cys	His	Ala	2896
2947	DB	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	2947
280	QY	Ser	Leu	Leu	Leu	Asn	Lys	Arg	Tyr	Leu	Gln	Leu	Tyr	Pro	-----	292
3001	DB	TG	TTT	ACCT	TTT	CGC	AAA	ACG	ACT	CG	AT	TTT	TAT	CTC	TTT	3060
293	QY	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	309
3061	DB	TC	GA	TAT	AT	ACG	GGG	CT	TA	ACG	CT	TA	CA	AA	CT	3120
309	QY	SG	Lys	Phe	His	Asp	Asp	Asp	Ala	His	Ala	His	Asn	Gly	Lys	329
3121	DB	TG	GT	TTT	TTT	TAAG	CGG	GGT	AT	CA	TTT	TCT	GA	CA	CA	3179
329	QY	Phe	Leu	Asn	Lys	Arg	Tyr	Glu	Leu	Arg	Ala	Glu	Trp	Lys	Gln	349
3180	DB	ATT	CAC	AC	AC	AC	AAA	AA	TTT	GC	AT	CCG	-----	-----	-----	3228
349	QY	u	Ala	Leu	Arg	Val	His	Leu	Asn	Arg	Asn	Asp	Tyr	His	Asp	369
3229	DB	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	3257
369	QY	a	Val	Glu	Asn	Phe	Asn	Asn	Gln	Thr	Gln	Asn	Ala	-----	-----	387
3258	DB	CAG	TGC	TCT	TTT	AT	CC	AA	GA	CA	GA	TAT	TAT	GT	GC	3309
387	QY	SG	In	Pro	Leu	Gly	Arg	-----	-----	-----	-----	-----	-----	-----	-----	393
3310	DB	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	3362
393	QY	u	Lys	Gly	Ser	Trp	Gly	Val	Gln	Tyr	Leu	Gly	Gln	Lys	Ser	413
3363	DB	AAA	AC	AG	CG	TGT	AGG	TAT	TGA	TAT	TTT	AC	GAA	ATA	AG	422
413	QY	RG	u	Ala	Val	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	428
3423	DB	CAA	AG	CAG	TGT	TAA	GTG	CTA	AT	CA	CA	AAA	CA	CA	TAT	3482
429	QY	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	429
3483	DB	GC	ATT	CG	GT	TTT	AT	CT	CTA	AT	CC	AA	GT	AA	GT	3542
429	QY	r	Ser	Phe	Phe	Gly	Val	Glu	-----	-----	-----	-----	-----	-----	-----	435
3543	DB	TT	CA	TAT	CA	TCT	CG	TAT	GAG	AA	TG	TTT	ATA	AG	AAA	3602
436	QY	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	439
3603	DB	TTT	AG	AAAA	AAAA	AA	TT	CA	AAAA	TT	GG	CTT	TACT	CA	TAT	3662

```

3723 GGCAGAGTATTTC-----GAGAGAGCTAATGAAACAGAGAAATGGTTACAA 3773
QY 471 sGln-----ProLeuProAspLeuGly-----478
Db 3774 AAAACAACCTTACTTATACCCAAAACCAACAGTAGGTTTGTAGTACAGATCATTTGTGA 3833
QY 479 -----AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAs 492
Db 3834 TTATAAAGTAACCTCTTAATTCAGAGACTGTAAAGTCGGTTAATTAAGGGAATAA 3893
QY 492 nTrpTyrPheThrProGlnHisLysLeuSerLeu-----503
Db 3894 TTATTATTTCGCGACGACGCAATAATATGCAATAGGGAATAACGTTGATTTAGGTTTAGG 3953
QY 504 -----ThrAlaSerHisGlnGlu-----509
Db 3954 TATTCGGTATGAGTATCTCGCACAAAGCTAATGAACTAATAGTGTGGTAAAT 4013
QY 509 -----509
Db 4014 TAAAAATTTCTCTTGGATACCTGTTATGTCATAAAACCAACGGAATGGCTTGATCTTTC 4073
QY 510 -----ArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLys 522
Db 4074 TTATCGCCTTCTACTGGAATTTAGAAATCCTAGTTTTCCTGAAATGTAT-----GGTTG 4127
QY 522 sHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAs 542
Db 4128 GCGGTATGTCGCAATAATAGCGAGGTTATGTAGTAAATTTAAAGCTGGAACATCTCG 4187
QY 542 nAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTy 562
Db 4188 TAACCAAGAGTTTGGTCTCGCTCTAAAGAGGGATTTTGGTAATATTTGAGATCAGTCATT 4247
QY 562 xArgAsnArgPheGlyAsnTyrIle---TyrAlaGlnThrLeuAsn-----AspG1 578
Db 4248 TAGTATGCTTATCGAAATCTTATCCCTTCTGAGAACTTAATAAAATGGAACCTGG 4307
QY 578 yArgGlyProLysSerIleGluAspSerGluMetLysLeuValArgTyrAsn---G1 597
Db 4308 AAAGGCCAATTTGATGATATCATATAATGCACAAATGCAAAATTAGTTGGCGTAATAATAC 4367
QY 597 nSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe-----611
Db 4368 TCGCAATTAGATTTTAAATGGTTTATGGAACGTTATCCCTACGGTTGGTGTGATGCAACATT 4427
QY 611 -----611
Db 4428 TGCTTATAACCGAGTAAAGATTAAAGATCAAAATAATCAATGCTGGTTTGGCCTCCGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 524
Db 4488 CAGTTATTTATTGATGCAATTCAGCCACG---CGTTATATCATTTGGTTTAGGCTATGA 4544
QY 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCAAGTAATACTTTGGGAATTAATAATGTTTCTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAs 653
Db 4605 AAATGAATTGCTAGGAATAACGT---GCATTTGGTAAACAAT-----TCAAGGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAGAAACTTACTCGGCA-----TGGCATATC-----4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTATGATGATCGGGTTATTATCATGCGCAATAAA-----4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713

```

4723 -----AATATTATGCTTCGATTAGGATATATATTTATTCAA 4760
QY 713 nThrArgTyrGlyGlnTrpAsnTrpTyrValLys---AlaAspAsnLeuLeuAsn-----730
Db 4761 CTATCGCTATGTTACTTTGGGAAGCGGTGCTCAACAGCACAGAGTCCGGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATATCTCGTACGACGATCA-----GGACGAAACTA 4868
QY 750 eThrGlyGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTAGAAATGAATTC 4893

RESULT 33
US-09-252-991A-14238
; Sequence 14238, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14238
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14238

Alignment Scores:
Pred. No.: 1,36e-08 Length: 2271
Score: 180.00 Matches: 179
Percent Similarity: 33.22% Conservative: 106
Best Local Similarity: 20.86% Mismatches: 276
Query Match: 4.46% Indels: 297
DB: 4 Gaps: 44

US-09-936-377-2 (1-759) x US-09-252-991A-14238 (1-2271)
QY 8 ProfileValLeuSerIleLeuLeu-----IleAsnThrProLeuLeuAlaGlnAla 24
Db 73 CCGCTGGCGCTTCCCTGCTCCTCTCGCTTCGTTTCCGCTCCTCGCTGCTGCT---GCC 129
QY 25 HisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgPro 44
Db 130 GATCCGGTGGAGCAGCAG-----ATGGTGGTATCGGCTCCGCGCGCGCG 174
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAsp 64
Db 175 ACCAGGATCAGCGAGTTGCCCGCAGCGCTCGGTGGTATCGAGCGC---GACCACTGGAC 231
QY 65 ThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGly---ValProGlyLle 83
Db 232 CAGCAGACCCAGCGCGCGCTGCCGCTGAAGGAGCGGTTGGGACAAATTGATCCCGCGCGCTG 291
QY 84 HisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArg 103
Db 292 GATATCGGCTCCAGGGTCTGTACCAACAGCGCGCAGAACCTTCGCGC-----GGACGC 342
QY 104 ArgIleLysValLeu-----AsnHisGlyGluThrGlyAspMetAlaAsp 119
Db 343 AGCGTACTGTGTGATGACGACGCGGTGTGCTGAACAGATTCCACGGGCGCATCAGCGCGCAG 402
QY 120 PheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeu 139
Db 403 TTCGAC-----TCCATCGACCCATCAACTCGAGCGCATCGAGGTGATG 447

US-09-252-991A-5404
 / Sequence 5404, Application US/09252991A
 / Patent No. 6551795
 / GENERAL INFORMATION:
 / APPLICANT: Marc J. Rubenfield et al.
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 / TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 / FILE REFERENCE: 107196.136
 / CURRENT APPLICATION NUMBER: US/09/252,991A
 / CURRENT FILING DATE: 1999-02-18
 / PRIOR APPLICATION NUMBER: US 60/074,788
 / PRIOR FILING DATE: 1998-02-18
 / PRIOR APPLICATION NUMBER: US 60/094,190
 / PRIOR FILING DATE: 1998-07-27
 / NUMBER OF SEQ ID NOS: 33142
 / SEQ ID NO 5404
 / LENGTH: 2403
 / TYPE: DNA
 / ORGANISM: Pseudomonas aeruginosa
 / US-09-252-991A-5404

Alignment Scores:
 Pred. No.: 1,49e-08 Length: 2403
 Score: 180.00 Matches: 173
 Percent Similarity: 37.02% Conservative: 103
 Best Local Similarity: 23.45% Mismatches: 264
 Query Match: 4.46% Indels: 218
 DB: 4 Gaps: 43

US-09-936-377-2 (1-758) x US-09-252-991A-5404 (1-2403)

Qy	41	LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle	60
Db	434	AAATCACCCTGCGC-ACCAGCAGCAGCTGTCTCCGCCACGCGCAGTCACCCGTC	492
Qy	61	IleSerGly-----AspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeu	77
Db	493	ATCGAGCATGCCCACTGAGGAGCTGCGCAGGGTCCGACAGCTGCCACCGTGCTG	552
Qy	78	---AspGlyValProGlyIleHisAlaSer-----GlnTyrglyGlyGly	91
Db	553	GCCAAAGCGGTGCGCGGATGTCGATTCAGCGCGCACCACCATCCGGAATAC-----	603
Qy	92	AlaSerAlaProValIleArgGlyGlnThr-----GlyArgGlnLysValLeuAsn	109
Db	604	-----GGCAGACCTCGCGCGGCGCAGCATGCTGGTGTGATGGTC	642
Qy	110	HisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAsp	129
Db	643	GACGGCGTGGCTTGAACACCAACCGCGACTCTCTCGCGCAACCTGGCC---AACATCGAC	699
Qy	130	ThrAlaLeuSerGlnGlnValGluLeuLeuArgGlyProValThrLeuLeuTyrrSerSer	149
Db	700	CCGGCGCTGATCGAGCGGATGAGGTATCTCGTGGCAGC---AGCGCCATCTACGGCAGC	756
Qy	150	GlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu	169
Db	757	GGCGCCCGCGCGGATCATCTCCATCACC-----ACCGTCCGCGCC	798
Qy	170	AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn---LeuGluLysLeu	188
Db	799	GGCGGC---GAGAACCGCGGAAACCCGCTCAGCGCCACCTCGCGCTGACCCGCGCTG	855
Qy	189	ThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu	208
Db	856	GGCAGCAT-----GGCTCGCGCGCCAGTTTC-----	882
Qy	209	TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAsp	228
Db	883	-----CAGCAATA-CTTCGCCGCTCCCTGGG	908
Qy	229	SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe	248

Db	909	GGCGCTCGACTATTCTTTCGACTTCGGCACCGCGTCA-----CGTGGCGGCTTC	956
Qy	249	TyrArgArgThrTyrrSerAspArgArg-----AspGlnTyrglyLeuProAlaHis-Se	266
Db	957	CTACGAGCCCATGGCAGCAGCATGCTCCCGGAACCCAGCAGCGACCTGTTCGACTC	1016
Qy	266	rHisGlnTyrr-----AspAspCysHisAlaAspI1	276
Db	1017	GAAGCTCTACAACTCGCGGCAAGCTCGGCTCGCATCGACGAG-----	1062
Qy	276	eIleTrpGlnLysSerLeuIleAsnLysArgTyrrLeuGlnLeuTyrrProHisLeuLeuTh	296
Db	1063	-----AACCGCGCTCGAGCTCGCCCTCAGCCACTACGACGC	1100
Qy	296	rGluGluAspValAspTyrrAspAsn-----ProGlyLeuSe	308
Db	1101	CGCCAGGACACCGACTACGCCACCGAGCGGTGCGCAGGCTGCGCGCGGC-----	1155
Qy	308	rCysGlyPheHisAspAspAlaHisAlaHisAsnGlyLysProTrpI1	328
Db	1156	-----TCGCTCCCGCCCAACCGCATCAAGGCGCTGGAG---CT	1190
Qy	328	eAspLeuArgAsnLysArgTyrrGluLeuArgAlaGluTrpLysGlnProPheProGlyPh	348
Db	1191	GGACGAGCGAACCCTATCCCAACACCTCGCGCAACCTCGAG-----TACGAGAACCT	1244
Qy	348	eGluAlaLeuArgValHisLeuAsnArgAsnAspTyrrHisHisAsp-----	363
Db	1245	CGACATCTCGCGCAGCGGCTCTCGCGCAGCTCTACTACCGTGACTATTTACCCGCTT	1304
Qy	364	-----GluLysAlaGlyAspAlaValGluAsnPhePh	374
Db	1305	CACCTCGTTTCAGCCCGCGCTCTCCACCGCGGCGCAACCTCGACAGATCATGCA	1364
Qy	374	eAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLy	394
Db	1365	GAACGAGCAAGTGTTCGCGCAGCGCTGACCTCGC---ACGCGCTGGCGGAAGCGG	1421
Qy	394	sGlySer-----TrpGlyValGlnTyrrLeuGlyGlnLysSer-----	406
Db	1422	CAATACCGAAGTGTCTGGGCGCGGACTACACAGAGACGACGACATCGCGCTCGA	1481
Qy	407	-----SerAlaLeuSerAlaThrSerGluAlaVal-----Ly	417
Db	1482	CGTGTTCAGCCCGCGGCTAGCAGCCAGCGCGGCTGTCTTCGACAGATCGGCA	1541
Qy	417	sGlnProMetLeuLeuAspAsnLysValGlnHisTyrrSerPheGlyValGluGlnAl	437
Db	1542	GCTCACCTACATGCGCGCTAGCAGCCCGCGCGCGCGCTCGGCCAGTTCGACGA	1601
Qy	437	aaasnTrpAsp---AsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSe	456
Db	1602	TCGCTTCGACGAACTGTGTGATCGAGCGGCGCTCGCTACGAATACTCCACGCGCGA	1661
Qy	456	rIleArgTyrrAspLysAlaLeuIle-----AspArgGluAs	468
Db	1662	ATTCGAGACTTCGCTCCGCTCTCCGAGTCAAGCGCGCTCGCGGTGACCGTCAAGG	1720
Qy	468	nTyrrTyrrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAl	488
Db	1721	G-----CGCGGACTCGACTACGAGCGGCTGTGCGAACCCTCGGCATCTCTATTCCG	1774
Qy	488	aLeuSerGlyAsnTrpTyrrPheThrGlnHisLysLeuSerLeuThrAlaSerHisG1	508
Db	1775	C-----GGTGCTGG-----CCAGGAATCTATGCTCTCTT-----CA	1807
Qy	508	nGluArgLeuProSerThrGlnGluLeuTyrrAlaHisGlyLysHisValAlaThr-AsnT	528
Db	1808	GCCAGGGCTTCAGCT-----GCCGATGTGGCATCCAGCTCGGCAACG	1852
Qy	528	hr-----PheGluValGlyAsnLysHisLeuAsnLysGluArgSerHisnAsnIleG	545
Db	1853	CCCGCGGTGGCTTCGATATCGGCTCTCTCGAACCTCGAACCGGTGAAGACCAACTACG	1912

```

Qy 545 lufeuAla---LeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyr-----562
Db 1913 AACTCGGCTGGCGGCCATCGGCGGCACACCCCTGGCGACCCCTGGCGCTGTTCATCA 1972
Qy 563 --ArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProL 582
Db 1973 CCACCTCAAGCTGGGCGAC-----GTGAGAGCTTCACACACGCG-----2013
Qy 582 ysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspP 602
Db 2014 -----CTGATCCTCCACCGCCACCAAGGAGCG-----A 2041
Qy 602 heTyrGlyAlaGluGlyGluIle---TyrPheLysProThrProArgTyrArgIleGlyV 621
Db 2042 TCTACGGCTCGAGCGCAGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2101
Qy 621 alSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluA 641
Db 2102 GCAGCGCCACCTGGATGCTGCGCGACGCGAGAAG-----CCGACGCGCAAGG 2146
Qy 641 spAlaTyrGlyAsnArgPropheIleAlaGlnAspAspGlnAsnAlaProArgValProA 661
Db 2147 AC-----TGGCAGGACATGACCGGCTACCGGCTGCGCGC 2179
Qy 661 laAlaArgLeuGlyPheHis-----LeuL 669
Db 2180 CGCTGAAGCTGACCGCCTACCTGCAATACAAAGCGGATGCGACTGGAAACACCGCCTGC 2239
Qy 669 ysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnA 689
Db 2240 AGGCACCTCTTCGACTCCAGGACTACCTGCTCGAGCGGCTGGAAAGCTTCGCGCGCG 2299
Qy 689 snLysLeuAlaArgTyrGluThrArgThrProGly-----HisHis 702
Db 2300 GCGAGTGCAGCAGCTACACACCGGT--CGACCTGCTGAGCGCAGTACCGCATCAC 2351

RESULT 35
US-09-307-973A-2
; Sequence 2, Application US/09307973A
; Patent No. 6432686
; GENERAL INFORMATION:
; APPLICANT: BULTHIUS, BEN
; APPLICANT: GATENBY, ANTHONY A.
; APPLICANT: TRIMBUR, DONALD E.
; APPLICANT: WHITED, GREGORY
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF
; TITLE OF INVENTION: 1,3-PROPANEDIOL BY RECOMBINANT
; TITLE OF INVENTION: ORGANISMS COMPRISING GENES FOR
; TITLE OF INVENTION: VITAMIN B12 TRANSPORT
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 925 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: U.S.A.
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT OFFICE 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/307,973A
; FILING DATE:
; CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,190
; FILING DATE: JUNE 30, 1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1245
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1844 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-307-973A-2

Alignment Scores:
Pred. No.: 2,52e-08 Length: 1844
Score: 176.00 Matches: 139
Percent Similarity: 34.31% Conservative: 83
Best Local Similarity: 21.48% Mismatches: 240
Query Match: 4.36% Indels: 186
DB: Gaps: 33

US-09-936-377-2 (1-758) x US-09-307-973A-2 (1-1844)
Qy 29 GlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSer 48
Db 61 CAGGACACTAGCCCGGATACCTGCTGCTCAGCCCAACCGCTTTTCAGCAGCGCGCAGC 120
Qy 49 GlyLeuLeuHisThrSerThrAlaSerAspLysLlelleSerGlyAspThrLeuArgGln 68
Db 121 GCGGTTCTGGCGCCCGTTCACCATC-----GTGACGCTGAGATATTGAACGCTGG 171
Qy 69 LysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyr 88
Db 172 CAATGACCTCGTAAATGATGTTCTGCGCGCTTTGCTGCGCTGATATTGGCAGAGC 231
Qy 89 GlyGlyGlyAla-----SerAlaProValIleArgGlyGlnThrGlyArgArgIleLys 106
Db 232 GCGCGCGCGCAGCAAAACTCCTCATTTTCATTCGCGCACCAACTCCAGCAGCATGCTAG 291
Qy 107 Val-----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
Db 292 GTATTGATTGACGCGCTGCTGATTTAGCAGCGCTGAGCGGTCC--GCCGATCTC 348
Qy 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArg 140
Db 349 AGC-----CAGTCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 393
Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAsp 160
Db 394 GGTCCGCGCTCGCTATTATTGTTCCGATGCTATCCGCGCGCGCTAGTGAATATC-----447
Qy 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuLeuLeuArg 180
Db 448 -----ATTACGCGCGCGAT 462
Qy 181 SerSerGlyAsnLeuGluLysLeuThrSerGly-----GlyIleAsnIleGlyLeuGlyLys 199
Db 463 AACCCAGGCA-----GAATTAACCGCTGATGGGGAAGCAAT-----AGTACCAGC 510
Qy 200 AsnPheValLeuHisThrGluGlyLeu-----208
Db 511 AATTACGACATCTCGACGCAACAGCAACTTGGCGAAATCACGCGCGCGCGCTGATCGGC 570
Qy 209 -----TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225

```

```

Db      571 GATTACGAATACACCAAGCGTTTACGGTGGTAGCGAAAGCGGACCGGATCGAGCG 630
QY      :|||||
Db      226 LeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArg 245
QY      :|||||
Db      631 CAGCTTGACCGGACGGCTTTTGTAGTAAACGCCTTATGGCGCGTTA----- 678
QY      :|||||
Db      246 LysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHis 265
QY      :|||||
Db      679 -----GAGCATACCTTTCTGATCCTGGAGCGGATTC-----GTGCGT 717
QY      :|||||
Db      266 SerHisGlyTyrAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLys 285
QY      :|||||
Db      718 GGTATGGCTACGATAAC----- 735
Db      286 ArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluAspValAspTyrAspAsnPro 305
QY      :|||||
Db      736 -----CGTACCGATTACGAC----- 750
QY      306 GlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLys 325
QY      :|||||
Db      751 -----GCCTATTACTCGCGCGGCTCG 771
QY      326 ProTyrPheLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPhe 345
QY      :|||||
Db      772 CCGCTGATTGATACACGC-----AAACTTTATAGCCAAAGCTGGGACCGCGGCTGCACATT 828
QY      :|||||
Db      346 ProGlyPheGluAlaLeuArgValHisLeu-----AsnArgAsnAspTyr 360
QY      :|||||
Db      829 AATGCG-----GAAAGTATTCACTCAGCTCGTGTTCAGCTATAGCCACAGCAATGATTAC 885
QY      :|||||
Db      361 HisHisAspGlnLysAlaGly-----AspAlaValGluAsnPhePheAsnGlnThr 378
QY      :|||||
Db      886 AACTATGATCCGCACTATGCGCGGTATGATACCTCCGCCACGCTGATGA-GATGAACA 944
QY      :|||||
Db      379 GlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGly 398
QY      :|||||
Db      945 GTACAATGTTCAATGGAC-----CAACAGTGTGTCGTGGGACCGTAACTGTGGG 995
QY      :|||||
Db      399 ValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGln 418
QY      :|||||
Db      996 GCGGCGGTAGACTGGCAGAAACAGACTACCACGCCAGGTACCGGCTATGTC----- 1046
QY      :|||||
Db      419 ProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn 438
QY      :|||||
Db      1047 CCGGAGGATATGACCAAGCGTATACCGGGTTTACCTGACAGGATTACACAGTTGGT 1106
QY      :|||||
Db      439 TrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArg 458
QY      :|||||
Db      1107 -----GACTTCATCTCTGGAGCGCGCGCGGCGGAGTGCATGACAACTCC----- 1148
QY      :|||||
Db      459 TyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGly 478
QY      :|||||
Db      1149 -----CAGTTTGGT 1157
QY      479 AlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGln 498
QY      :|||||
Db      1158 CGTCAT-----GGTACATGCGCAACACCGCGCGGATGGAGTTTATAGAAGT 1205
QY      :|||||
Db      499 HisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyr 518
QY      :|||||
Db      1206 TATCGCTTTATGCTCTCTAGGAACTCTACAAAGCGCTAATTTGGCCAACTGTAT 1265
QY      :|||||
Db      519 AlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLys 538
QY      :|||||
Db      1266 GGTATT-----TACGTTAATCCGAACCTGAATCCT 1295
QY      :|||||
Db      539 GluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu-----GlyAspArg 554
QY      :|||||
Db      1296 GAAAGAGTAAACAGTGGAGGCGCA-----TTTGAAGGCTAACCGCTGCGTCACG 1349
QY      :|||||
Db      555 TrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThr 574
QY      :|||||

```

```

Db      1350 TGGCGT-----ATTTCAGTTATCGTAAAC-----GAT 1376
QY      575 LeuAsnAspGlyArgGlyProLysSerIleGluAspSerGluMetLysLeuValArg 594
QY      :|||||
Db      1377 ATTAATGAC-----ATGATCGATTATGACGATCATCTGCATAAATAT--- 1418
QY      :|||||
Db      595 TyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThr 614
QY      :|||||
Db      1419 TACACGAGAGTNAAGCGCGCATTAAGTATTAGGCGACCGCGAATTCGATACCGGA 1478
QY      :|||||
Db      615 ProArgTyrArgIleGlyValSerGlyAspTyrVal-----ArgGlyArgLeuLysAsn 632
QY      :|||||
Db      1479 CCGTTAACGCATACG---GTCAGTTATGATTACGTTGATGCGCGTAATCGGATTACCGAT 1535
QY      :|||||
Db      633 LeuProSerLeuProGlyArg 639
QY      :|||||
Db      1536 ACGCCA---TTACCCCGCGGT 1553
QY      :|||||
RESULT 36
US-09-307-973A-1
; Sequence 1, Application US/09307973A
; Patent No. 6432686
; GENERAL INFORMATION:
; APPLICANT: BULPHIUS, BEN
; APPLICANT: GATENBY, ANTHONY A.
; APPLICANT: TRIMBUR, DONALD E.
; APPLICANT: WHITED, GREGORY
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF
; TITLE OF INVENTION: 1,3-PROPANEDIOL BY RECOMBINANT
; TITLE OF INVENTION: ORGANISMS COMPRISING GENES FOR
; TITLE OF INVENTION: VITAMIN B12 TRANSPORT
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.
; STREET: 925 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: U.S.A.
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT OFFICE 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/307,973A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,190
; FILING DATE: JUNE 30, 1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1245
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

```

ANTI-SENSE: NO
US-09-307-973A-1

Alignment Scores:

Pred. No.: 3,21e-08 Length: 1845
Score: 175.00 Matches: 131
Percent Similarity: 31.87% Conservative: 80
Best Local Similarity: 19.79% Mismatches: 236
Query Match: 4.34% Indels: 235
DB: 4 Gaps: 30

US-09-936-377-2 (1-758) x US-09-307-973A-1 (1-1845)

QY 29 GlnSerValGlyLeuGlnThrValThrValValGlyLys-----SerArgProArgAla 46
DB 61 CAGGATACCGCGGATACCTCGTCTACTGCTAACCGTTTGAACAGCGCGCAGC 120
QY 47 ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleleSerGlyAspThrLeu 66
DB 121 ACT-----GTGCTTGACCAACACCGCTT-----GTGACCGCTCAGGATATCGAC 165
QY 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
DB 166 CGCTGGCAGTCGCTCGTCAATGATGTGCTGCGCGCTCTCCGGCGGCGATATCACC 225
QY 87 GlnTyrGlyGlyAla-----SerAlaProValIleArgGlyGlnThrGlyArg 104
DB 226 CAAAACGGCGCTCAGGTCAGCTCTCATCTATTTTATTCGCGGTACAAATGCCAGTCAT 285
QY 105 IleLysVal-----LeuAsnHisGlyGlnThrGlyAspMetAla 118
DB 286 GTGTGGGTAAATGATGGCGTACGCTGAATCTGGCGGGGAGTGTTCT---GCC 342
QY 119 AspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGlu 138
DB 343 GACCTTAGC-----CAGTTCCTATTGCGCTTGTCAGCGGTGTGATAT 367
QY 139 LeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspVal 158
DB 388 ATCCGTCGGCGCGCTCGCTGTTATGTTCCGATGCAATAGCGGGGGTGGTAAATATC 447
QY 159 AlaAspGlyIleProGluLysMetProGluAsnGlyValSerGlyGluLeu 178
DB 448 ATCAGCGCGC-----GATGAACCGGACG 474
QY 179 ArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly 198
DB 475 GAAATT-----TCAGAGGGGTGGGAGCAATAGTTAT 507
QY 199 LysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValPro 218
DB 508 CAGAACTATGATCTCTACGAG-----531
QY 219 ArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHis 238
DB 532 -----CAACAACCTGGGGATAGACACGCGGTAAACGCTGTTGGCGGATAT 576
QY 239 ArgAlaValLeuGlyTyrArgLysArgPheTyrArgThrTyrSerAspArgArgAsp 258
DB 576 -----576
QY 259 GlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleTyr 278
DB 577 -----GCCATACCTCATGTTATGAT-----597
QY 279 GlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGlu 298
DB 598 -----GTTGTGCTATGATGATATACCGGACCGACGACG 630
QY 299 AspValAspTyrAspAsnProGlyLeuSerCysGlyPhe-----311
DB 631 CAGACAGATACGAT-----GGTTTTTTTAAAGTAAACGCTTTATGCG 672

QY 312 -----HisAspAspAspAla-----317
DB 673 GCGCTGAGCATATCTTTACTGATGCTGAGCGGCTTTGTGCGGCTATGCTATGAT 732
QY 318 -----HisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArg---331
DB 733 AACCGTACCAATATGACGGGTATTTCTCCGGTTCACCGTGTGCTGATACCGGTAA 792
QY 332 -----AsnLysArgTyrGlyLeuArgAlaGluTrpLysGlnProPheProGlyPheGlu 349
DB 793 CTCTATAGCCAAAGTTGGACGCGCGCTGCTATAAACGCGCACTGATTAATACACA 852
QY 350 AlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGly-----367
DB 853 CTCATTACCACTATACCATAGCAAGATTAACACTACGATCCCATATATGCTGTTAT 912
QY 368 -----AspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAla 381
DB 913 GATTCTGCGCGACGCTCGATGATGAAGCAATACACCGCTCCAGTGGGCAACAAT---969
QY 382 ArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyr 401
DB 970 ---GTATCTGTTGTCAGGTAGTATTGGTGG-----GGTGTGACTGG 1011
QY 402 LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeu 421
DB 1012 -----CAGAAACAGACTACGACCGCGGTACAGT-----1041
QY 422 LeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrp-----439
DB 1042 -----TATGTTGAGGATGGATATGATCAACGTAATACCGGCATC 1080
QY 440 -----AspAsnPheThrLeuGluGlyGlyValGluValGlu 451
DB 1081 TATCTGACCGGCTGCAACAGTCGGCGATTTCCTTTGAAGCGCCAGCAGTGAC 1140
QY 452 LysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys 471
DB 1141 GATAACTCA-----1149
QY 472 GlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGly 491
DB 1150 -----CAGTTTGGTCTGCTAT-----GGAACCTGCAACACCGCGCC 1185
QY 492 AsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnLysArgLeu 511
DB 1186 GGTTCGGAATTCATCGAAGGTTATTCGCTTCATTCCTACCGGACATCTTATAAGGCA 1245
QY 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
DB 1246 CCAATCTGGGCACTGTATGGCTC-----TAC 1275
QY 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
DB 1276 GGAATTCGGAATCTGGACCGGAGAAACACAGTGGGAGGCGG-----TTTGAA 1329
QY 552 -----GlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGly 567
DB 1330 GGTTCACCGCTGGGCTGAACTGGCT-----ATTCCGATATCTAACGATGTCAGT 1383
QY 568 AsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAsp 587
DB 1384 GACTTGATC-----GATTAT 1398
QY 588 SerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGly 607
DB 1399 GATGATCACCGCTGAAATATTAACAGAGGAAAGCGGATTAAGCGCTGAGCGC 1458
QY 608 GluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArg 627
DB 1459 ACCGCAATTTTGTATCCGACCACTGACGCTACT---GTGACTTATGATTATGTCAT 1515
QY 628 GlyArg 629

```

Db      1516 GCGCGC 1521
RESULT 37
US-09-328-352-1816
; Sequence 1816, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1816
; LENGTH: 2286
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1816

Alignment Scores:
Pred. No.: 5,79e-08 Length: 2286
Score: 174.00 Matches: 181
Percent Similarity: 32.86% Conservative: 117
Best Local Similarity: 19.96% Mismatches: 299
Query Match: 4.31% Indels: 310
DB: 4 Gaps: 46

US-09-936-377-2 (1-758) x US-09-328-352-1816 (1-2286)
QY 7 LysProIleValLeuSerIleLeu-----LeuIleAsnThrProLeuLeuAla 22
Db 28 AAGCGTATTATTCAGTCAGTGTCTTCTTTCAGTACTGCGAAGCATGATGCTATGGCT 87
QY 23 GlnAlaHisGluThrGluGlnSer----- 30
Db 88 TTTGCTGCACAAATGAGCAAGATCAAGCTGAAAGAAACATTAGAAAAGCGCTGACCT 147
QY 31 ValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGlyLeu 50
Db 148 GTGAATTTGGAACAAATTTTCGTACACCTGAAGAGCAAGTGAAGCAATCGCTCGGTGTA 207
QY 51 LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla 70
Db 208 -----TCGCTTATTACCAAGAGATTTAGAAAAAATACCA 243
QY 71 Val-----AsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyr 88
Db 244 GTCTGAATGATATTCTGACTATGTCGTATGTCGATGCGAGGTGTCACACCTGACA----- 297
QY 89 GlyGlyAlaSerAlaProValIleArgGlyGlnThr----- 101
Db 298 -----GGCAATAGTGTACAGGCGAGCGTGTGTAATATAGACAAATGATATTCGCGGA 351
QY 102 -----GlyArgGlyIleLysValLeuAsn--- 109
Db 352 ATGGGCGCTGAAACACGCTTATTTTAGTGACGCTAAACCGATTAATCTCTGTAATTC 411
QY 110 --HisHisGlyGluThrGlyAspMet-----AlaAspPheSerProAsp 123
Db 412 GTTCGTATGCTGCGAAGAGAGCGGTATACACGAGCGGACTCAAACTGGGTACACGA 471
QY 124 HisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
Db 472 GAAGCAATC-----GAGTCGATCGAAGATTTTTCGTTGACCGACCA 510
QY 144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIle 163
Db 511 GCAGCTCGTATGGCTCTGGTCTCGGGTGTGTAGTTACATCATCTACTAAAAAGTA 570
QY 164 -----ProGlyLysMetProGlu 169
      |||||

```

```

Db      571 ACBAATGAACACTCATGTTTCAGTAGAGTCTATATCTTCACAGCCTGAAGACTCCAAAGAA 630
QY 170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSer----- 182
Db 631 ---GGTTCATCAAAATCGTGTGGTGTAAATGTAAGTGGCCCATTAATTAAGAGAGCTTTTG 687
QY 183 -----GlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIle----- 195
Db 688 TCTTATCGTTTATATGTTAATCAATAAACGGAAGCTGATGATGATGATATATAAA 747
QY 196 GlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys-----SerGly 213
Db 748 TCTATCGGTAGTACA---GCAGCTGGCGGTGAAGGTGTAAATAATAAGATATTTCCAGC 804
QY 214 -----AspTyrAlaValProArg 219
Db 805 CGTTTAGCTTGGCAGGCAACAGACCAAACTCTTCTGCGATATTTCTTCTAGCAAA 864
QY 220 TyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg 239
Db 865 CAAGGTAATATT-----TATTCTGGTCACTCTCAGTTAAATGCAAAATGCTGAAGCGGAT 918
QY 240 AlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGln 259
Db 919 GCGATTCITTTCA-----CAGCTGATGCTGAAGAAACCAATCCTATGATGATGATGATG 972
QY 260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGln 279
Db 973 TATGCATTA-----ACGCAGAA-----GGTGATTGGTCTTGGGT 1008
QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLysAsp 299
Db 1009 AAGAGTAAGTATGTTGCTCAATATGATGAAGCCATCAACAACTGCTACCTGAA----- 1062
QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAsp-----Asp 314
Db 1063 -----GGCTTGGCGGAAGTGTAGAAAGGAAAAATTAATAATCTTGTAT 1104
QY 315 AspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuAsnLysArg 334
Db 1105 GATAAAGCCACTCG-----CGTTTGAAGAACT 1131
QY 335 TyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPhe-----GluAlaLeuArg 352
Db 1132 CTTTCGCTTAAACGCGAGGCTAAATATCTCTTTGATATCTATTTACCCCAAGTATTAAC 1191
QY 353 ValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsn 372
Db 1192 GTA-----GGTACCGAATGGGTTGAAGAC 1215
QY 373 PhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArg 392
Db 1216 AGATTTAAAGATAATGCTCGACAACT-----CAA 1245
QY 393 LeuLysGlySerTrpGlyValGlnTyr-----LeuGlyGlnLysSerSer 407
Db 1246 GGTAAAGACAGCAGCGGTTCAGGTTTCAGGATCAATTTAGCGAAGGTTGATCGTAGTAA 1305
QY 408 AlaLeuSerAlaThrSerGluAlaVal-----LysGln 418
Db 1306 ATGGAGTCACTGCTTCTGCTATATATTGAGATAACCTGAAAGTTACAGACGACCA 1365
QY 419 ProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn 438
Db 1366 GATGTTGATTAGGTTTACGTTTTCAGTATGATGATTTTCACTTTAAAGGTGGG 1416
QY 439 Trp-----AspAsnPheThrLeuGluGly 447
Db 1417 TGGAGCCCAAGCTTAAATATTACTCAAAACTCAATGATTTTCACTTTAAAGGTGGG 1476
QY 448 ValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGlu 467
Db 1477 GTAGCA-----AAAGCT----- 1488

```


QY 468 AsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaSerPhe 487
 Db 1489 -----TATAAGCACCA-----AATATGTATCAAAATGCCGAAGGTAT 1527
 QY 488 AlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis 507
 Db 1528 TATTAAAGTACAAATGGCAATGGCTGTCT----- 1557
 QY 508 GlnGluArgLeuProSerThrGlnGlnLeuTyrAlaHisGlyLysHisValAlaThrAsn 527
 Db 1558 -----GCTAATATTGAGTCGGT 1575
 QY 528 ThrPheGluValGlyAsnLysHisLysAsnLysGluArgSerAsnAsnIleGluLeuAla 547
 Db 1576 TGTTTATTACAGGTAAATGGTAACTGAAACATCGTAACAAAGAGCTCGT 1635
 QY 548 LeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGly 567
 Db 1636 ATTCAAGTCCAAAGAGATATCGTAATCGAGCTTAACCTGCTCCGTAATGATATAAA 1695
 QY 568 AsnTyrIleTyrAlaGlnThr-----LeuAsnAspGlyArgGlyProLysSer 583
 Db 1696 GATAAGATTGTCGGGTACTCATGTTGCGAACAGTTGATGGCTCAAGTACAAATGCA 1755
 QY 584 -----IleGluAspAspSerGluMetLysLeuValArgTyr---AsnGlnSerGly 599
 Db 1756 AATACAGGAGCTGTGACCAATACGAAGTGAATATTTGCGTTGGAAATAACGCTAAA 1815
 QY 600 AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIle 619
 Db 1816 GCCTTAATTCAGGTTTTCAGAGAGT-----TTG 1845
 QY 620 GlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArg 639
 Db 1846 GGGTTAGACTTCGGTGATATCGCTGG---ACTAATAACTTTACCTACATGATGACTCG 1902
 QY 640 GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgVal 659
 Db 1903 AAAGACAAGCAACACGGGAACCATTA-----TCTTTAGTT 1938
 QY 660 ProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
 Db 1939 CCAATCTATACAATTAATCAATTTTGTGATTATGACATCACTGATCAATTCGATGATAAT 1998
 QY 680 LeuAspTyr-----TyrArgValPheAlaGlnAsnLysLeu--- 691
 Db 1999 TTTGTTTACTCAATATGCTCGTGCATAATCAGCTCAATTTGCAGAAATAGACTGAA 2058
 QY 692 -----AlaArgTyr 694
 Db 2059 TCCGGTATAGTTTCAGGAGTGCGAATTCCTGCGTTAAGCCAAAGTACGTAATAAGTTAT 2118
 QY 695 GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyr-----ArgArg 712
 Db 2119 AGTACCGCTGGT-----ATTAAAGTTGGTTATAAGTTTCAGACCAAAAT 2163
 QY 713 AsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAlaAspAsnLeuAsnGlnSer 732
 Db 2164 AGTACAGCTGTGGT-----GTGAGTAATCTGTTTGATAAACA 2202
 QY 733 ValTyrAlaHisSerPheLeuSerAspThr---ProGlnMetGlyArgSerPheThr 751
 Db 2203 ATTTTAAAGACAGATTAATCTATTAGCCAACTATATATGACCCAGGTCGAGCTTATTAC 2262
 QY 752 GlyGlyValAsnValLysPhe 758
 Db 2263 GCATCTTTAAATAATATCTTTC 2283

RESULT 38

US-09-453-702B-174

; Sequence 174, Application US/09453702B

; Patent No. 6365723

GENERAL INFORMATION:
 APPLICANT: Blattner, Frederick R.
 Burland, Valerie
 Perna, Nicole T.
 Plunkett, Guy
 Welch, Rod
 TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
 NUMBER OF SEQUENCES: 265
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Quarles & Brady
 STREET: 1 South Pinckney Street
 CITY: Madison
 STATE: WI
 COUNTRY: US
 ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 8.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/453,702B
 FILING DATE: 03-Dec-1999
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/110,955
 FILING DATE: 04-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27396
 REFERENCE/DOCKET NUMBER: 960296.95017
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (608) 251-5000
 TELEFAX: (608) 251-9166
 INFORMATION FOR SEQ ID NO: 174:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7304
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 174:
 US-09-453-702B-174
 Alignment Scores:
 Pred. No.: 3,89e-07 Length: 7304
 Score: 174.00 Matches: 171
 Percent Similarity: 31.41% Conservative: 101
 Best Local Similarity: 19.75% Mismatches: 288
 Query Match: 4.31% Indels: 306
 DB: 4 Gaps: 42
 US-09-936-377-2 (1-758) x US-09-453-702B-174 (1-7304)
 QY 1 MetAlaGlnThrThr-----LeuLysProIleValLeuSerIleLeuLeu 15
 Db 5095 ATGGCTAAGTTCACACTTCCTCAGGAATCAAGTTCGGCGCTCTTTTCACTGCTC 5154
 QY 16 IleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThr 35
 Db 5155 TTTCGACGACCAATGATT---CATGCAACCGACACTGCACGACCAAGATGGCGAAACA 5211
 QY 36 ValThrValVal-----GlyLysSerArgProArgAlaThrSerGlyLeu-----Ileu 51
 Db 5212 ATCATCTGTTTACAGCGGATGCAAAATACCGCACTGAGCGGACCGATGTTTATCAACTCTG 5271
 QY 52 HisThrSerThrAla----- 56
 Db 5272 AGCACTTCCACGCGGACATTAACCGATATGCGATGCTGGATATCCCGAGGTGCTCAAT 5331
 QY 57 -----SerAspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGly 74
 Db 5332 ACGGTTAGCGATCAGGTTCTCGAA-----AATCAGAATGCAACGACGCTGGAT 5379

QY 75 AspaLeuAspGlyValProGly---lleHisAlaSerGlnTyrGlyGlyAlaSer 93
 Db : : : : :
 5380 GAAGCGCTTTATACGTCAGTAACGTTGGTACAGACCAATACATTAGGCGGAACTCAGGAC 5439
 QY 94 AlaProValIleArgGlnThrGlyArgArgGlyLeuValLeuAsnHisHisGlyGlu 113
 Db : : : : :
 5440 GCCTTTGACGTCGTTGTTGGTCTAACCGG----- 5472
 QY 114 ThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAsp----- 129
 Db : : : : :
 5473 -----GATGGCTCCATCATGACCAACCGCTCGCAACT 5505
 QY 130 -----ThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
 Db : : : : :
 5506 GTACTTCCTCGCAGTTTCAACCGCCGACAGACGCTGTGGAAGTCTAAAGGTCGGCC 5565
 QY 144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIle 163
 Db : : : : :
 5566 TCCACGCTGTATGGCATTCGATCTCGTGGTGAATTGATTAACTCGTGACC---AAGCGC 5622
 QY 164 ProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly 183
 Db : : : : :
 5623 CCGGAAAAACATTCATGGTCTGCTCAGCC----- 5655
 QY 184 AsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLys----- 199
 Db : : : : :
 5656 -----ACGTCCTCCAGTTTGGCGAGGACCTGGGCAACTTGATATCACA 5700
 QY 200 AsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArg 219
 Db : : : : :
 5701 GTCCCATTCAGGCACTCAGTGGCATACCGCTGACGGGGAAGTG----- 5748
 QY 220 TyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg 239
 Db : : : : :
 5748 ----- 5748
 QY 240 AlaValLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGln 259
 Db : : : : :
 5749 -----CAGATGAAGATTACTGGCGAAACTTCGGTAAAGACGCGAGTACA 5793
 QY 260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
 Db : : : : :
 5794 TTTATTGGCCGCTCA-----CTCAGCTGGTT 5820
 QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHis----- 293
 Db : : : : :
 5821 GGTGAT-----ATGCAACAGTAACCATGCTCTATTCCATCGGAGCTATAAATCCG 5874
 QY 294 -----LeuLeuThrGluGluAspValAsp----- 301
 Db : : : : :
 5875 TTCGATCGTGGACGATTTTCGACCTTACGAGCAACACCGCTAAACGTTGATCGAANA 5934
 QY 302 -----TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
 Db : : : : :
 5935 ATACGTTTGGACGACCG-----TTTAAATTACAGAT----- 5967
 QY 320 HisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAla 339
 Db : : : : :
 5968 -----GTCAGTCCGATCTGGCGCAACTCAACGCGAGATATCATCTCAATAGC 6015
 QY 340 GluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
 Db : : : : :
 6016 CAGTGGACGCGGCTTT-----GAT 6036
 QY 360 TyrHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGln----- 377
 Db : : : : :
 6037 TACAGCTACAGCCAG-----GATAAATACGCGACATCAGGCTCGC 6078
 QY 378 -----ThrGlnAsnAlaArgIleGluLeu----- 385
 Db : : : : :
 6079 GTTACCGCGTATGATGCAACGACGAGCACTGACACGCGCTGTTGATGCAACTCAGGGA 6138

QY 386 -----ArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyr 401
 Db : : : : :
 6139 TTTACCCACGCGTATGCAATCTACTCGTGGGATCTGCAAGGGAAT-----GTTGATATT 6192
 QY 402 LeuGlyGlnLysSerSerAlaLeuSerAlaThrSer-----GluAlaValLys 417
 Db : : : : :
 6193 GCTGGGTTCTATAATAGATTTCTGGGTGGGTGTCTATGAATATATGATCTTCTCGCGC 6252
 QY 418 GlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAla 437
 Db : : : : :
 6253 ACAGATATATCGCTGTGTAAACCGTAAGATTTCAATATCTACACACCCGTTTATGGC 6312
 QY 438 AsnTrpAspAsnPheThrLeu-----GluGlyGlyValArgValGluLysGln 453
 Db : : : : :
 6313 AATACCAGCAATGCAACACGTTTCGGCTCGGACAGTGTACAGCATCAACACAGGAG 6372
 QY 454 LysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsn----- 468
 Db : : : : :
 6373 AGCTACTCAGCTTATGACACAGGATCGCTCTATCTGACCGATTAATGGATTGCCGTCGCC 6432
 QY 469 -----TyrTyr----- 470
 Db : : : : :
 6433 GGGATCCGCTATCAGTATTACACGCGATGTCGGGTAAAGCCGCTCTTTTAAATGTCAT 6492
 QY 471 -----LysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArg 485
 Db : : : : :
 6493 ACTGACAGCGCGATGCAACATGCAACGCGCAAACTGGGGTTAGTCTACAAACTGACGCCA 6552
 QY 486 SerPheAlaLeuSerGlyAsnTrp-----TyrPheThrProGlnHisLysLeuSerLeu 503
 Db : : : : :
 6553 TCGGTATCTTATTTGCTAAATTTATTCGCAACATTTATGCGCAA-----TCGTCA 6603
 QY 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
 Db : : : : :
 6604 ATTGCCAGCTACATCGAGATCTCCACCGGAA----- 6636
 QY 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
 Db : : : : :
 6637 ---TCATCTAAATGCTTACCAAGTCGGGGCAAAA----- 6666
 QY 544 IleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArg 563
 Db : : : : :
 6667 TTCGAGCTGTTCGATGCTATAACCGCAGAT-----ATTGGCTGTTCAT 6711
 QY 564 AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp----- 577
 Db : : : : :
 6712 ATCCATAACGTAACTGTTGTATACCGAAAGTATTTGGTATGAAACCATCCGCAAAACG 6771
 QY 578 ---GlyArg---GlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr 595
 Db : : : : :
 6772 GCAGGCGCGGTTGCTTCAAGAGGGGTAGAAATCGACCTTGGCGGAGCATTAATCTGAAAC 6831
 QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
 Db : : : : :
 6832 ATTAATATCATTCGACGATACCGCTATACCGATGCTAAGGTTCTGGAGATCCT----- 6885
 QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysLeuProSer 635
 Db : : : : :
 6886 -----GATTATGCGAGGAAACCATTCGCGAATGTTCTCTCGT 6921
 QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspGlnAsn 655
 Db : : : : :
 6922 CATACCGGT-----TCGCTATTCTCGTACCTATGACATTCATTAAC 6960
 QY 656 AlaProArgValProAlaAlaArgLeu-----GlyPheHisLeu-----LysAlaSer 671
 Db : : : : :
 6961 ATCCAGGCAATAACACACTGACGTTGGCGGTGGTGCNATTCGCTAAGCGCTGCTCG 7020
 QY 672 LeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeu 691
 Db : : : : :
 7021 GCAACCAAT-----GGGGTGTACTATTAT----- 7044
 QY 692 AlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711

Db 7045 -----CTGCCAGGCTATTTCTGGCGATGCTCTCGCCGATACACAAA 7086
 Qy 712 ArgAsnThrArgTyrGlyGluTrpAsnTyrValLysAlaAspAsnLeuAsnGln 731
 Db 7087 ATGAATTGAGTAT---CCGGTCACACTGCAATTAAACGTCAAAAACCTGTTTGATAA 7143
 Qy 732 SerValTyrAlaHisSer 737
 Db 7144 ACGTATTACACCTCTTCC 7161
 RESULT 39
 US-09-252-991A-1708
 ; Sequence 1708, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 1708
 ; LENGTH: 2172
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-1708

Alignment Scores:

Pred. No.: 6,01e-08 Length: 2172
 Score: 173.50 Matches: 181
 Percent Similarity: 33.89% Conservative: 103
 Best Local Similarity: 21.60% Mismatches: 285
 Query Match: 4.30% Indels: 272
 DB: 4 Gaps: 44

US-09-936-377-2 (1-758) x US-09-252-991A-1708 (1-2172)

Qy 7 LysProIleValLeuSerIle-----LeuLeuIleAsnThrProLeuLeuAlaGln 23
 Db 40 GAGCCTTCGCTCGCCCTCCACCGCTCTGCTGTACCGCGCCCTCGCCGAA 99
 Qy 24 AlaHisGluThrGluGlnSerValGlyLeuGluThrVal 38
 Db 100 ACCGAGGAGCGCCCTCGCCCTCGCCCGAGCCAGTGTGTCAGCGTTGCGCAGGATCCCGCC 159
 Qy 39 -----ValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThrSer 54
 Db 160 GAACCTCGACCATCGACCTCGCCACTCCGGTCAGTCCGGTTCGGCTTCGGCTTCAGC 219
 Qy 55 Thr-----AlaSerAspLysIleIleSerGlyAspThrLeuArg 67
 Db 220 GCCCTGGACACCCCGCGGAGCAGCAGCATCAGCGCGGAGGAGTTCGCGGAGCGCAAC 279
 Qy 68 -----GlnLysAlaValAsn-----LeuGlyAspAla 76
 Db 280 AACCCAGCGTCGAGCGCGCTCCACCGCAGCCCGGATCATCTCATCGGCCACCCA 339
 Qy 77 LeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal 96
 Db 340 GGATGCGCGGTACCGCGCTTCGCGCGCGGTTCAGCGCCGCGCAGCATCG-----GTG 393
 Qy 97 IleArgGlyGlnThrGlyArgGlyLeuValLeuAsnHisGlyGlyGluThrGlyAsp 116
 Db 394 ATGCACTGTCGAGCGGACCGGCTGTACACCGG-CATGGGACCGGTGAAGCTT-----446
 Qy 117 MetAlaAspPheSerProAspHisAlaIle-MetValAspThrAlaLeuSerGlnGlnVa 136
 Db 117 MetAlaAspPheSerProAspHisAlaIle-MetValAspThrAlaLeuSerGlnGlnVa 136

Db 447 -----CCCAGGCGATCCGTGGATGGTFC-----GAGCGCAT 476
 Qy 136 lGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVa 156
 Db 477 CGACGTGATCCGGCGCCCGCTCGTGTACGGCGAAGCGCCACCGCGCGGTGAT 536
 Qy 156 lAspValAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLe 176
 Db 537 CAACGTG-----GTGCGAAGAACGCTTTCGCGCGGAGAT 572
 Qy 176 u-----GlyLeuArgLeuSerSerGlyAsnLeuGluLys-----LeuThrSe 190
 Db 573 CCGCAACCACTCGCTCGCTACGGCTCTACGACCAACCGCCAGCTTGGCCCTGGACAG 632
 Qy 190 rGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu---Ty 209
 Db 633 CGCGCGC-----TCGCTGACCGACAGCAGCTCAGCTA 662
 Qy 209 rArgLysSerGlyAspTyrAlaVal---ProArgTyrArgAsnLeuLysArgLeuProAs 228
 Db 663 CGGCTCAACCT-GAACCAGCAGCAGCAGCGCTG-----GA 700
 Qy 228 pSerProArgArgPheAlaAsnGlyGlnHisArg-AlaValLeuGlyTyrArgLysArg- 247
 Db 701 TCACCGTGGCGACTCGCGCAACCTCGGCGATCAGCGCGCGCTGCGCTGCGAGGCGAGG 760
 Qy 248 -----PheTyrArgArgThrTyrSerAspArgArg-----AspGlnTyrG 261
 Db 761 ACGATCTGGCTTCACCTCGCCACGACTATGGCGACCAGGAGCGGATGAACGACTTCG 820
 Qy 261 lLeuProAlaHisSerHisGlyTyrAspCysHisAlaSerIleIleTrpGlnLys 281
 Db 821 GCACCCCGCTGCTCGCGCGCAAGTAC-----CACAAGC 853
 Qy 281 erLeuIleAsnLysArgTyr-----LeuGlnLeuTyrProHisLeuLeuT 296
 Db 854 GCCTGCGCGAGAGAACTACAACGTCGCAACGAGCTGCGAGCGCTACACGACCACTGGA 913
 Qy 296 hrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAs 316
 Db 914 CCGCTCTGACCGAGCGCTG-----AGCCTTTCC-----G 943
 Qy 316 spAlaHisAlaHisAlaHisAsnGlyLysProIleAspLeuArgAsnLysArgTyrG 336
 Db 944 ACAGC-----GTACCCGCGAGCAACAGTCTACT 973
 Qy 336 lueLeuArgAla-----GluTyrLysGlnProPheP 346
 Db 974 ACATCAAGGCGCGCGCTGCGCAACGCGCAACCTACGAATGGGACGCTCCCG----- 1029
 Qy 346 roGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr-----HisH 362
 Db 1030 -----CGCGAAGAGCTGTTCGCGAGGAGTACCTGCGCATCAGCCAGC 1072
 Qy 362 isAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaA 382
 Db 1073 AGCAGGAGCAGATCGCGCAGCCGACAGCTTCGCTTC-----CAGCAGCGCC 1120
 Qy 382 rgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrL 402
 Db 1121 T-GTTCGCTCGACCGCCACCTGTCGCGCGCA----- 1157
 Qy 402 euGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuL 422
 Db 1158 -----GTACAACCGCATCCGC 1173
 Qy 422 euAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsn 442
 Db 1174 TTCGCTGACCAACACATATCGCCCTATAC----- 1205
 Qy 442 heThrLeuGluGlyGlyValValGluLysGlnLysAlaSerIleArgTyrAspLysA 462
 Db 1206 -----CGATGTCGCGCGCGACTACATCGAC 1230

```
QY 462 laLeuileaspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArg 481
Db 1231 CCTGGCATCCGGCGCGGCTACTTGAAGCGGTTCGCCCTACCGGCGGCACTCGCGC 1290
QY 482 GlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeu 501
Db 1291 AGCCAGCCCGCACTGCTCGCCCTGTCGCGGAGAACCGCTGCAGCTCAACGAGCGCTG 1350
QY 502 Ser-----LeuThr 504
Db 1351 TCGTGTGACCGGGGTGCGCGCGACAGCAACCATATACCGCGACGACCTGCTCAC 1410
QY 505 AlaSerHisGlnGluArg-----LeuPro 512
Db 1411 GGCACCCGCGACCGACGACCTGCGAGCGGCACTGGCGCGCGGATGCTTCGCG 1470
QY 513 SerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn-----527
Db 1471 CTGACCCCGGAAGTGTGCTGTACGCGCCAGTACTCCACCGAGCGAGCGGTGTGAGCAAC 1530
QY 528 -----ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGlu 545
Db 1531 CTGATCACCTCAACCGCGCGCCGCGAGATGGACCTGACCCACCTCGAAGCAGACCGAG 1590
QY 546 LeuAlaLeuGly-----TyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArg 563
Db 1591 GTCGGCTCAAGCAGCTGTTCGCGAGCGTTCGCGGCGGATGACCTGCGCGCTAC---1647
QY 564 AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSer 583
Db 1648 -----CATATGCTCAAGAAAGAGTCTCAGCGCC---AACCGCTGCGC 1689
QY 584 IleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db 1690 CCGCAGCAGCGCAGCAGTTC-----GGCCAGCAGAGTTCGAGC-----1728
QY 604 GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db 1729 GGCCTGGAGCGCAGCTGGAACCTGAACCTGGCCAGGACTGGCGGCTGTTCGCGAGCGC 1788
QY 624 AspTyrValArgLysLeuLysAsnLeu-----ProSerLeuProGlyArgGlu---Asp 641
Db 1789 GCGTGGTGTGCGCGAATACGACGACTTCGACGAAACGATGACGCGGACGACCTACTCG 1848
QY 642 AlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
Db 1849 CGCAACGGCAATCGCCC-----AGGAAGTGCACCG---CGCAC 1887
QY 662 AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
Db 1888 GCCAACCTCTGGCTGGACAAAGTCTTCGCGGAGACCTCGCGGTGCGCGCGGCTGCGC 1947
QY 682 TyrTyr---ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
Db 1948 TACGTGACCGCGCTACGCGAC-----GGCGCAACACGCGCCAGCCTGCGCGCGC 1998
QY 701 HisHisMetLeu-----AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGlu 718
Db 1999 TACACGGTGTGCTCCAACTCGGC-----2025
QY 719 TrpAsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyr 734
Db 2026 -----TGCGGGTACGGCGGACCTGACCTCGCGCTTGAGCTGTAT 2067
```

RESULT 40
US-09-252-991A-2016/c
; Sequence 2016, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 2016  
; LENGTH: 2241  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-2016  
  
Alignment Scores:  
Pred. No.: 6.32e-08 Length: 2241  
Score: 173.50 Matches: 181  
Percent Similarity: 33.89% Conservative: 103  
Best Local Similarity: 21.60% Mismatches: 285  
Query Match: 4.30% Indels: 272  
DB: 4 Gaps: 44  
  
US-09-936-377-2 (1-758) x US-09-252-991A-2016 (1-2241)  
  
QY 7 LysProIleValLeuSerIle-----LeuLeuIleAsnThrProLeuLeuAlaGln 23  
Db 2196 GAGCCTTCGCGCTCGGCTCTCCACAGCCTGCTGTGTACCGCCGCGCTCGCGAA 2137  
  
QY 24 AlaHisGluThrGluGlnSerValGlyLeuGluThrValThrVal-----38  
Db 2136 ACCGAGGAAGCGCCCTCGCCCTGCCAGCCAGTGTGTAGCGTTCGCCAGGATCCGCC 2077  
  
QY 39 -----ValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThrSer 54  
Db 2076 GAACCTCGACCATACGACCTCGCCACTCGGTGTCAGTCCGCTTCGCGCTCGCGCTCAGC 2017  
  
QY 55 Thr-----AlaSerAspLysIleIleSerGlyAspThrLeuArg-----67  
Db 2016 GGCCTGGACACCGCGCGAGCACACGATCATCAGCGCGAGAGTTCGCGACGCAAC 1957  
  
QY 68 -----GlnLysAlaValAsn-----LeuGlyAspAla 76  
Db 1956 AACCCAGCGTCCAGCGCGCGCTCACCCGACGCCCGGATCATGCTTCATCGCACCCCA 1897  
  
QY 77 LeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProVal 96  
Db 1896 GGAGATGGCGGTACCGCGCTCTCGCGCGCGCTTCAGCGCGCGCCACGATCG-----GTG 1843  
  
QY 97 IleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAsp 116  
Db 1842 ATGCAACTGTTTCGACGCGACGCGCTGTACACCGG-CATGGGACGCGTGAACCT-----1790  
  
QY 117 MetAlaAspPheSerProAspHisAlaIle-MetValAspThrAlaLeuSerGlnGlnVa 136  
Db 1789 -----CCCAGCGATCCGTGTGATGTC-----GAGCGCAT 1760  
  
QY 136 IclulleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVa 156  
Db 1759 CGACGTGATCGCGCGCGCGCTCGGTGTGTACGCGAAGCGCCACCGCGCGCGTGTAT 1700  
  
QY 156 laspValAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLe 176  
Db 1699 CMACTG-----GTGCCGAAGAACCTTTTCGCGCGGAGAT 1664  
  
QY 176 u-----GlyLeuArgLeuSerSerGlyAsnLeuGluLys-----LeuThrSe 190  
Db 1663 CGCAACCCACTCGCGCTCGGTCTCTACGACACCGCCAGTGTGGCCCTGGACAG 1604  
  
QY 190 rGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu---Ty 209  
Db 1603 CGCGCGC-----TCGCTGACCGACGACGCTCAGCTA 1574  
  
QY 209 rArgLysSerGlyAspTyrAlaVal---ProArgTyrArgAsnLeuLysArgLeuProAs 228
```

```

Db      1573  CCGGCTCAAGCT-GAACCAGCAGCAGCAGCCAGCGCTG-----GA 1536
Qy      228  pSerProArgArgPheAlaAsnGlyGlnHisArg-AlaValLeuGlyTrpArgLysArg- 247
Db      1535  TCGACCTGGCGACTCGGCAACTGGGATCAGCGGGCGCTGCGCTGGCGCCAGCG 1476
Qy      248  -----PheTyrArgArgThrTyrSerArgArg-----AspGlnTyrG 261
Db      1475  ACGATCTGGCCTTCAACCTCGCCACAGCTATGGCAGCAGCAGCGATGAACAGCTTCG 1416
Qy      261  lyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAspIleIleTrpGlnLys 281
Db      1415  GCACCCGCTGGTGGCGGCAAGTAC-----CACAAAGC 1383
Qy      281  erLeuIleAsnLysArgTyr-----LeuGlnLeuTyrProHisLeuLeuT 296
Db      1382  GCCTGGCGGAGAAACTTACAACGTGCGCAACGAGTGCAGCGCTTACAACAGCAGTGA 1323
Qy      296  hrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspA 316
Db      1322  CCGCCTGACAGCAGCTGG-----AGCCTTTC-----G 1293
Qy      316  spAlaHisAlaHisAlaHisGlnGlyLysProTrpIleAspLeuArgAsnLysArgTyrG 336
Db      1292  ACAGC-----GTACCGCGGAGCAACCACTTGTACT 1263
Qy      336  luLeuArgAla-----GluTrpLysGlnProPheP 346
Db      1262  ACATCAAGCGCGCGCCACTGGCGCAACCGCAACCTACGATGGAGCTCCG----- 1207
Qy      346  roGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr-----HisH 362
Db      1206  -----CGCAAGAGCTGTGTGGCAGGACTACTTGGCGATCAGGCACG 1164
Qy      362  leAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAla 382
Db      1163  AGCAGAGCAGATCGGCGACCGCAGACCTTCGCTTC-----CAGCAGCC 1116
Qy      382  rgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrL 402
Db      1115  T-GTTCGGCTCGACAGCGCAGCCAGCTGTGTGGCGCCGA----- 1079
Qy      402  euGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuL 422
Db      1078  -----GTACACCGCATCCGC 1063
Qy      422  euAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAspAsnP 442
Db      1062  TTCCGCTTGACCAACACATATCGCCCTATAC----- 1031
Qy      442  heThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysA 462
Db      1030  -----CGATGTCGGCGCGACTACATCGAC 1006
Qy      462  laLeuIleAspArgGluAsnTyrTyrLysGlnProLeu-ProAspLeuGlyAlaHisArg 481
Db      1005  CCCTGGCATCGCGCGCGCTACTTCGAAAGCGCTGCGCTTACCGCGCGCATCGCGC 946
Qy      482  GlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeu 501
Db      945  AGCCAGACCGGACCTTCGCTCTGTCGCCGAGAACCGCTGCGAGCTCAACGAGCGCCTG 886
Qy      502  Ser-----LeuThr 504
Db      885  TCGCTGGTGACCGGGGTGGCGCGACCAACCATATCGACCGCGAGCAGCTGCTCACC 826
Qy      505  AlaSerHisGlnGluArg-----LeuPro 512
Db      825  GGCACCGCGAGCGCGAGCGCTGCGAGCGGCGCACTGCGCGCGGATGGTCTTCGCG 766
Qy      513  SerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn----- 527

```

```

Db      765  CTGACCCCGGAACTGTGCTGTACGGCCACTCTCCACAGCAGCAGCAGCGGTGTGACCAAC 706
Qy      528  -----ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGlu 545
Db      705  CTGATCACCTCAACGGCGGCCAGCAGCATGACCTGACCCACTCGAAGCAGACCGAG 646
Qy      546  LeuAlaLeuGly-----TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg 563
Db      645  GTCGCGCTCAAGCAGCTGTTCCCGGAGCGTCCGCGGAGTGGACCCCTGGCGCGCTAC--- 589
Qy      564  AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSer 583
Db      588  -----CATATCGTCAAGAGAAAGCTGCTCAGCGCC---AACCCGCTGCCG 547
Qy      584  IleGluAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db      546  CCGCACACCGCGCAGCAGGTC-----GGCCAGCAGAGTTCGGAC----- 508
Qy      604  GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db      507  GGCCTGGAGGCCAGCTTGAACCTGCGCCAGGACTGGCGCTGTGCGGGAACGCC 448
Qy      624  AspTyrValArgGlyArgLeuLysAsnLeu---ProSerLeuProGlyArgGlu---Asp 641
Db      447  GCGCTGTGTGCTGCCAATACGACGACTTCGACGAAACGATCGACGCGGCACTACTCG 388
Qy      642  AlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAla 661
Db      387  CGCAACGGCAATCGCCCC-----AGGAACGTGCCACCG---CGCACC 349
Qy      662  AlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
Db      348  GCCAACCTCTGCTGGGCAAGTCTCTCGCGAGACCTGCGCGCTGCGCGCGCTGCGC 289
Qy      682  TyrTyr---ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGly 700
Db      288  TAGCTCGACCGCGCTACGCGAC-----GCCGCCAACCGAGCGCGCTGCGCGC 238
Qy      701  HisHisMetLeu-----AsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGlu 718
Db      237  TACACGGTGGTTCGACTCCACACTGGGC----- 211
Qy      719  TrpAsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyr 734
Db      210  -----TGGCGGTACGGCGGACCTGACCTCGGCTCGGCTGAGCTGTAC 169

```

Search completed: November 17, 2003, 19:58:13
Job time : 1837 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 19:22:57 ; Search time 461 Seconds
(without alignments)
5373.750 Million cell updates/sec

Title: US-09-936-377-2
Perfect score: 4036
Sequence: 1 MAQTLKPIVLILLINTEL.....FLSDTFQMGSRFTGTVNVKF 758

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0936377/runat_14112003.104451.22224/app.query.fasta.1.903
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0936377@cgn_1.154@runat_14112003.104451.22224
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3928	97.3	2295	14	US-10-066-551-7

Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-10-066-551-7
; Sequence 7, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; TITLE OF INVENTION: prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045US1
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US/10/066,551
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 7
; LENGTH: 2295
; TYPE: DNA

Sequence 1, Appli
Sequence 8, Appli
Sequence 3, Appli
Sequence 54, Appli
Sequence 1, Appli
Sequence 194, App
Sequence 731, App
Sequence 172, App
Sequence 507, App
Sequence 11, Appl
Sequence 526, App
Sequence 4, Appli
Sequence 4030, Ap
Sequence 262, App
Sequence 152, App
Sequence 261, App
Sequence 24, Appl
Sequence 174, App
Sequence 13, Appl
Sequence 16, Appl
Sequence 109, App
Sequence 3, Appli
Sequence 2, Appli
Sequence 119, App
Sequence 719, App
Sequence 59, Appl
Sequence 1, Appli
Sequence 720, App
Sequence 57, Appl
Sequence 1, Appli
Sequence 1076, Ap
Sequence 5, Appli
Sequence 1049, Ap
Sequence 1, Appli
Sequence 105, App
Sequence 1, Appli
Sequence 3, Appli
Sequence 6035, Ap
Sequence 333, App
Sequence 1, Appli
Sequence 4051, Ap
Sequence 284, App


```
Db 1998 GGCCTGGCGCTCGGCTCCACTGAAGCCCTGCTGACGCGCATGATGCCAATT 2057
Qy 680 uAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProG 700
Db 2058 GGAATACTACCGGTGTTCGCCCAAAACAACTGCGCCGCTACGAAACGCGACGCCGG 2117
Qy 700 yHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTyrAs 720
Db 2118 ACACCATATGCTCAACTCGGCGCAAACTACCGCCCAATACGCGTATGGGAGTGGAA 2177
Qy 720 nTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPhe 740
Db 2178 TTGGTACGTCAAAACGCGACAACTGCTCAACCAATCCGTTTACGCCCAACAGAGCTTCT 2237
Qy 740 uSerAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 2238 CTCTGATACGCCCGCAATGCGCGCAGCTTTACCGCGCGGTGAAGTTT 2292

RESULT 2
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; FILE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c

; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
```


QY 272 -----CysHisAlaSpIlelle-----TrpGlnLysSerLeu 282
Db 292555 ACTCTAAAGTGAACACACGTTAGGTAGTCTAGTATTATTAAGTGAATCAACAAATTCCTG 292614
QY 283 ILeAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyr 302
Db 292615 ATCAAAGCGGTATTAAAC-TATTAT-----TTAAGGCA----- 292649
QY 303 AspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHis 322
Db 292650 GATAATCTTATCTA-----AATACGCATATCGCACTGTAT 292685
QY 323 AsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLys 342
Db 292686 AACAAATAAACTATTGAGAAAGAACACGCGTAAGTCTAGTGGTGTGAAAGATCGACATAA 292745
QY 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsn----- 358
Db 292746 TTGACTACCGAGGT-----ATAAATTACGTAATTTCTCCGAATTATCT 292790
QY 359 -----AspTyrHisAspGluLysAlaGlyAspAlaVal 370
Db 292791 CACATTTCTTTGTTATGGGTGCATATATATCGAGATAAAATCCGTACCGAA----- 292844
QY 371 GluAsnPheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIle 390
Db 292845 -----CGAGGCACAAACGGTAGCGAGTGGAAAGTTCCGAGCGGACCCCTAT 292889
QY 391 GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSer 410
Db 292890 AATCGAATCAACACTACAGCGGT-----TATTTA----- 292922
QY 411 AlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSer 430
Db 292923 -----ATGCCCATATATTCG 292937
QY 431 PhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgVal 450
Db 292938 CTATTGGG----- 292946
QY 451 GluLys-----GlnLysAlaSerIleArgTyrAsp-----LysAla 462
Db 292947 GAAAAATGTAGTTTCGCAAGTGTAGTTATGACCACTACGATACCTCAAGTAACACC 293006
QY 463 LeuLeuAspArgGluAsnTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln 482
Db 293007 GTAAATAACAAGGATAATCATTTA----- 293030
QY 483 ThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer 502
Db 293031 -----TCTCTGCCACAAAATTAAGTGGATAGTAC-----AATGGTTAGAT 293075
QY 503 LeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHis 520
Db 293076 TTTACTGCCAAATATAATGAAGCTTTCGAGCACCATCTATGCAAGAGCGATTTGTGAGT 293135
QY 521 GlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsn 533
Db 293136 GGTGTCACCTTGGGCGCAAACTCTAGGCTAGATCATCAATAGATTTGTAGCAAAAT 293195
QY 534 LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeu----- 546
Db 293196 CCAATTTGGCCCTGAACAGCGGAAAATAAGAAATTAACGGAATCTACATTTTGTAT 293255
QY 547 AlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPhe 566
Db 293256 AGTCTGTTTAAACAAGCGGATAAAATCAAAATTTGAAGCGACTTATTTCGGAATATGATGTG 293315
QY 567 GlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp-----GlyArg 579
Db 293316 AAAGATTTTATTAACTTAAATAATTTAATGATGCAAAAGACAAAGTGCAGGTGCA 293375
QY 580 GlyProLysSer-----IleGluAspAspSerGluMetLysLeuValArgTyr 595

Db 293376 AATCCAAATCAAAATGAGCATTTGTCGCAAAAATTCACAGTATCAA----- 293423
QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyLeuTyrPheLysProThrPro 615
Db 293424 AACATAACTAATGCCCGTTTAAAGCGGTATTGAATTGCAAGCTCAATACCAA---ACAGAA 293480
QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
Db 293481 CGTTAACGGTA-----TTTACTAATCTATGGCAGC 293510
QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
Db 293511 ACCAAAGTAAAGATAAAGATAGTGGC-----GAAGCT 293543
QY 656 AlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAsp--- 674
Db 293544 TTATCAACATTTGCCCAAGCAAAATCGCGTACGGGTAAATTTATGTTAGTAAAGAC 293603
QY 675 -----ArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
Db 293604 AATTCACGGTGGGAGCGACAGTAACCCATTACGCT-----GCTCAACGCCGAGTGCCT 293657
QY 693 ArgTyrGluThrArgThr-----ProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
Db 293658 AAAGATCATAGTGTACTACCCAGTTATATATACTACGCGATCTTCGTGCTACCTATGCA 293717
QY 712 ArgAsnThrArgTyrGlyGluTrp 719
Db 293718 -----CCATTAAAGGCGAATGG 293735

RESULT 3
US-10-066-551-8
; Sequence 8, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045U1
; CURRENT APPLICATION NUMBER: US/10/066.551
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 8
; LENGTH: 2307
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-8

Alignment Scores:
Pred. No.: 2,44e-14 Length: 2307
Score: 227.50 Matches: 189
Percent Similarity: 33.14% Conservative: 101
Best Local Similarity: 21.60% Mismatches: 299
Query Match: 5.64% Indels: 287
DB: 14 Gaps: 43
US-09-936-377-2 (1-758) x US-10-066-551-8 (1-2307)
QY 16 IleAsnThrProLeu-----LeuAla 22
Db 99 ATGAATACCCCATTTGTTCCGCTCAGCCTGCTCGCTCACTTGGCGGAGTTTGGC 158

QY 23 GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer 42
 Db 159 CACGGCGCAGAAAATAATGCAAGGTGCTAGTACCGTACTGTGTAAGGCGACCGC 218
 QY 43 ArgProArgAlaThrSerGlyLeuLeuHisThrThrAlaSerAspLysIleIleSer 62
 Db 219 CAA-----GGCAGCAAAATCCGTACCAACATCGTT--- 248
 QY 63 GlyAspThrLeuArgGlnLys-----AlaValAsnLeuGlyAspAlaLeuAsp 78
 Db 249 -----ACGTGCAACAAAAGACGAAAGCAGCACCGATGATGCGCAACTTTAAA 302
 QY 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIle--- 97
 Db 303 GAAGAGCGTCCATC-----GATTCGGCGCGCAACGCGACGTCCCAATTCCGT 353
 QY 98 -----ArgGlyGlnThrGlyArgArgIleLysVal----- 107
 Db 354 ACGTCGCGCGCATGGGTGAGAACTCTGTGACATCAAGGTGGAACAGCGCTATTCCGAC 413
 QY 108 -----LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis 124
 Db 414 AGCCAAATCCTTTTACCACCAAGGCAGA----- 440
 QY 125 AlaIleMetValAspThrAlaLeuSerGlnValGluLeuLeuArgGlyProValThr 144
 Db 441 ---TTTATTGTCGATCCGCTTTGGTTAAAGTCGTTTCCGTACAAAAGGC----- 488
 QY 145 LeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro 164
 Db 489 -----CGGGTTCGCTCTCGCGGTATCGCGGACCAACGCGCGCATCATC 536
 QY 165 GluLysMetProGlu-----AsnGlyValSerGlyGluLeuGlyLeuArg 179
 Db 537 GCCAAAACCGTCGATGCCAAGACCTGCTCAAAGGCTTGGATAAAAAGTGGGCGTCGC 596
 QY 180 LeuSerSerGly-----AsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeu 197
 Db 597 CTCAACAGCGGCTTTGCGCAGCAAGGCGGTAAAGTACGCGCAAGCGTATTCCGAAAA 656
 QY 198 GlyLysAsnPheValLeuHisThrGluGlyLeuTyr-----ArgLysSerGly 213
 Db 657 GAGGGCAACTTC-----GACGGCTTGTTCTTACACCGCAACGATGAAAAA 704
 QY 214 AspTyrAlaValProArg-----TyrArgAsnLeu-----LysArgLeuProAspSer 229
 Db 705 GATTACGAAGCGCGCAAAAGGTTTCCGCAATGTCAACGCGCGCAAAACCGTACGACGC 764
 QY 230 Pro-----ArgargPheAlaAsnGlyGlnHis 238
 Db 765 GCGCTGGACAAACGACGTACTCGCAAAATCGGAACACCTTCGCGCAGCAGCAC 824
 QY 239 ArgAlaValLeuGlyTyrArgLysArgPheTyrArgArg----- 251
 Db 825 CGCATGTTGTGAGCCACATGAAAGACCAACACCGGGGCATCCGCACTGTGCGTGAAGAA 884
 QY 252 ---ThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAsp 270
 Db 885 TTTACCTCGCGCAAAAAGTTTACGGATAAATATTGACCGCCCAAGCCCTGCTTACCGC 944
 QY 271 AspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeu 290
 Db 945 GAA----- 947
 QY 291 TyrProHisLeuLeuThrGluLysValAspTyrAspAsnProGlyLeuSerCysGly 310
 Db 948 -----ACTACCAATCCACACCACTTGGGTATACGCGGTAAACCTCTGGC 995
 QY 311 PheHisAspAspAspAlaHisAlaHisAsnGlyLysProTyrPheAspLeu 330
 Db 996 TTTGTCGAAAAAATCGGATGCCAAGCCTATGTGTT---GGAAAAAGAACGCTATTCCGC 1051
 QY 331 ArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAla 350

Db 1052 CGATGACAGCGGACCGGCTTACGAGCAATGTAAAGGCCCAACCA----- 1099
 QY 351 LeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaVal 370
 Db 1099 ----- 1099
 QY 371 GluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro-I 390
 Db 1100 -----TACCCGAATCACCACTCGTGTGCGAACTTCAACTTCGACAGCGCCTGC 1150
 QY 390 eGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSe 410
 Db 1151 CGAACAAACCTCTTGAATACGCTATCACTACCGCCATCAGAA----- 1196
 QY 410 xAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSe 430
 Db 1197 -----ATCAACCGCGAAGCATTTTGAATCGAAA-----TTCTC 1231
 QY 430 rPhePheGlyValIleGluAlaSerIleArgTyrAspLysAlaLeuLeuAspArgGluAsnTyrTy 470
 Db 1232 CATCCCGACGACAGAGAGAAAAAC-----GGTCAAAAAGT 1267
 QY 450 lGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspArgGluAsnTyrTy 470
 Db 1268 CGATAAACCGATGGAAACAACAATGAAGACCGGTGCGAGATGAAGACACTGTTCACGCTA 1327
 QY 470 rLysGlnProLeuPro-----AspLeuGlyAlaHisArgGlnThrAlaArg----- 485
 Db 1328 CAACTTTCACACCGACCAACCGATACCGGTATATGTTGAAGCCATTACGACAT 1387
 QY 486 ---SerPheAlaLeuSerGlyAsn----- 492
 Db 1388 CGGCGATTTCACTGACGCGCGGCTGCTTACGACCGCTTCAAGGTGAAAAACCCATGA 1447
 QY 493 -----TrpTyrPheTh 496
 Db 1448 CGCAAAAACCGTTTCAAGCAGCAACCTTAACCGAGTTTCGGTGTGATTTGG-----CA 1501
 QY 496 rProGlnHisLysLeuSerLeuThrAlaSerHisGln-----GluArgLeuProSerTh 514
 Db 1502 GCGCAGCAACACTGGAGCTTCAGCGGAGCCCACTACGCCAGCGCGACCCCGCGCT 1561
 QY 514 rGlnGlu---LeuTyrAlaHisGlyLysHis-----ValAlaThrAsnThrPh 529
 Db 1562 GTATGACGCGCTGCAAAACCCACGCTAAACCGCGCATCATCTCGATTGCGACGCGACA-- 1619
 QY 529 eGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGl 549
 Db 1620 -----AAAGCGCAACGCGCGCAATACCGAAATCGGCTTCAA 1657
 QY 549 yTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
 Db 1658 CTACAACAGCGCGCAGTTTCCGCAAC-----GGCAGCTA 1693
 QY 569 rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerGl 589
 Db 1694 CTTCTGG---CAGACCATCAAGACGCGCTTGCATCCGCAAAACCGCCACGACTCTGT 1750
 QY 589 uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyGl 608
 Db 1751 CGCGCTCGGTGAGCCGCTCAATGCCGTTATCATCAAAAACCGCTTACGAAATGGGCGC 1810
 QY 608 uIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGl 628
 Db 1811 GTCTACCGCACCGCGCGCTGCTGCTCAAAAGTCGCGTCAGC----- 1853
 QY 628 yArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPh 648
 Db 1854 -----CACAGCAAAACCGCGCTTTACGATACGCAACAGCAAGCTGTT 1897
 QY 648 eIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLe 668

Db 1898 GAGCGCAATCTGAATTT-----GGGCAACAAGTCGGCGCACTTG 1939

Qy 668 uIysAlaSerLeuThrAspArgIle---AspAlaAsnLeuAspTyr---TyrArgValph 686

Db 1940 GAGCGGCTCCCTTGGCTACCGCTTCCAAATCGGAATCTGGAATCGGCTGGGCGC-----1994

Qy 686 eAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLe 706

Db 1995 -----GGCGGTATTGTTCAAAAAGCTACGGGT---TCGATATTGGCGGC 2035

Qy 706 uGlyAlaAsnTyrArgArg-----As 713

Db 2036 AGGTCAAAAGACCGCAAGCAACTTGGAAACGTTGTACGCAAGGTTTCGGTGTGAA 2095

Qy 713 nThrArgTyrGlyGluTyr-----AsnTyrTyrValIysAl 725

Db 2096 CGATGCTTTCGCCCACTGGAAACCGCTGGGCAAGACACACGCTCAATGTCAATCTTTCGGT 2155

Qy 725 aAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---Pr 744

Db 2156 TACACAGGTGTTCAACAGTCTTACTATTCGCAAGCAAGCTGGGCAAGCTGGACCAATCCCTGCC 2215

Qy 744 oGlnMetGlyArgSerPheThrGlyGlyValAsnValIysPhe 758

Db 2216 GGCGGTGGGACGTGATGATGCTGCGCTGGCGTGAACTACAAGTTC 2258

RESULT 4

US-09-801-451A-3

; Sequence 3, Application US/09801451A

; Publication No. US20030096369A1

; GENERAL INFORMATION:

; APPLICANT: Sparling, P. Frederick

; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule

; and Vaccine

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ImClone Systems Incorporated

; STREET: 160 Varick Street

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10014

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/801,451A

; FILING DATE: 08-Mar-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/418,964

; FILING DATE: 07-APR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Gallagher, Thomas C.

; REGISTRATION NUMBER: 37,066

; REFERENCE/DOCKET NUMBER: SPA-3-P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 645-1405

; TELEFAX: (212) 645-2054

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2381 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 172..2313

; SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-801-451A-3

Alignment Scores:

Score: 7,97e-14 Length: 2381

Score: 223.00 Matches: 193

Percent Similarity: 33.45% Conservatives: 99

Best Local Similarity: 22.11% Mismatches: 291

Query Match: 5.53% Indels: 292

DB: 11 Gaps: 45

US-09-936-377-2 (1-758) x US-09-801-451A-3 (1-2381)

Qy 11 LeuSerIleLeuLeuIleAsnThrProLeu---LeuAlaGlnAlaHisGluThrGluGln 29

Db 193 CTGAGCTGTCTCTCGCTCACACTTGGCGCGGCTTGGCCACGCGGAGAAAATAATGTC 252

Qy 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGly 49

Db 253 AATGTGCAATTGGATACCGTTACCGTAAAGAGCGACGCCAA-----294

Qy 50 LeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLys 69

Db 295 -----GGCAGCAAAAATCCGTACCACATCGTT-----ACGCTTCAACAAAA 336

Qy 70 -----AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla 85

Db 337 GACGAAAGCACCGCAACCGCATATGCGGAACTCTTAAAGAGAGCCCTCCATC-----390

Qy 86 SerGlnTyrGlyGlyAlaSerAlaProValIle-----ArgGlyGln 100

Db 391 ---GATTTGCGCGCGGCAACGCGCACGTCCCAATTCCTGACGCTGCGCGCATGGTCA 447

Qy 101 ThrGlyArgArgIleLysVal-----LeuAsnHis 111

Db 448 AACTCTGTGACATCAAGGTGGACAAACGCTATTTCGACAGCCAAATCTTTACCACAA 507

Qy 112 GlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAla 131

Db 508 GGCAGA-----TTTATTGTGATCCCGCT 531

Qy 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151

Db 532 TTGTTAAAGTCGTTTCCGTACAAAAGGC-----GCGGGTTCC 570

Qy 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu-----169

Db 571 GCCTCTCGCGGTATCGGCGCGCACCAACGCGCGGATTATCGCAAAACCGTCGATGCCAA 630

Qy 170 -----AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186

Db 631 GACCTGCTCAAGGCTTGGATAAAACTGGGGCGTGCGCTCAACAGCGGC-----681

Qy 187 LysLeuThrSerGlyGlyIleAsnIleGly-----LeuGlyLys-----AsnPhe 201

Db 682 TTTGCGGCAACAACGCGTAAGCTACGCGCAAGCGTATTTCGAAAAAGAGGCAACTTC 741

Qy 202 ValLeuHisThrGluGlyLeuTyr-----ArgLysSerGlyAspTyrAlaVal 217

Db 742 -----GACGGTTTGTCTCTTACAACCGCAACGATGAAAAAGATTACGAAGCC 789

Qy 218 ProArg---TyrArgAsnLeu-----LysArgLeuProAspSerPro-----230

Db 790 GGCAAGGCTTCCCAATGTCAACGCGGCAACCGTACCGTACGCGCGCTGGACAAA 849

Qy 231 -----ArgArgPheAlaAsnGlyGlnHisArgAlaValLeu 242

Db 850 CGCAGCTACCTCGCCAAAATCGGAACAACCTTCGCGAGCGCGACCGCATCGTATTG 909

Qy 243 GlyTyrArgLysArgPheTyrArg-----ArgThrTyrSerAspArgArgAspGlnTyr 260

Db 910 AGCCATATGAAGACCAACACCGCGGCGCATCGCACTGTG-----CGTGAAGAGTTT 960

Qy 261 GlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLys 280

```

Db 961 GCGTGGCGGCGGAAAT-----TCA 981
Qy 281 SerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspVal 300
Db 982 CGGATAACTATTAAACGCCAAGCCCTGCTACCGGAA---ACCACACAATCCAAACC 1038
Qy 301 AspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHis 320
Db 1039 AATTGGCGTACCGCGCAAGATTGGGCTTTGTCGAAAACTGGATGCCAAGCCTAT 1098
Qy 321 AlaHisAsnGlyLysProTyrPheLeuAspLeuAsnLysArgTyrGluLeuArgAlaGlu 340
Db 1099 GTGCT---GGAAAAAAGCGTATTCCGCGATGACAAAGATAACCGCTACGCGAGCAA 1154
Qy 341 TrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyr 360
Db 1155 TGTAAAGGCCCCNACCA-----TACCCGAATCGCCACTCGGG 1193
Qy 361 HisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsn 380
Db 1194 CAT----- 1196
Qy 381 AlaArgIleGluLeuArgHisGlnPro-IleGlyArgLeuLysGlySerTrpGlyValG 400
Db 1197 ---GAACCTTCAACTTCACAGCCGCTTGGCGAACAAACCCCTGTTGAAATACGGCATCAA 1253
Qy 400 nTyrLeuGlyGlnLys-----SerSe 407
Db 1254 CTACCGCATCAGGAATCAACCGCAAGCGTTTTTGAATTCACAATTTAAATTTGAAGA 1313
Qy 407 rAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValG 427
Db 1314 TAAAAAAGATGCAACTGAGGAAGATATAAAAGAAAGACCGTGAATAATGAAAAATGCGCAA 1373
Qy 427 nHisTyrSerPheGlyValGluGlnAlaAsnTrp----- 439
Db 1374 AGCTACCGTCTGACCAACCGCAGCAAAACCGATACCGCGCGGTATATCGAAGCCATTCA 1433
Qy 440 -----AspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerI 457
Db 1434 CGAGATTGACGGCTTTACCTGACCGCGGCTGGTTACGACCGCTTCAAGTGAAC 1493
Qy 457 eArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys---GlnProLeuProAs 476
Db 1494 CCAC-----GACGG-CAAAACCGTTTCAAGCAGCAGCAGCTCAACCGCA 1534
Qy 476 pLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheT 496
Db 1535 GTTT-----CGCGTGATTTGGCAGCCCGCGCGCTG-----TATGACG 1621
Qy 496 hrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnG 516
Db 1577 GCGGAGCCACAACCTAC-----GCCAGCGCAGCGCGCGCTG-----TATGACG 1621
Qy 516 luLeuTyrAlaHisGlyLysHis-----ValAlaThrAsnThrPheGluValG 532
Db 1622 CGCTGCAAAACCGGCAAAACCGCGCATCTCGATTGCGAGCGGCAAC----- 1671
Qy 532 lyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluG 552
Db 1672 -----AAGCCGAACCGCGCGCAATACCGGAATCGGCTTCACTACAACG 1717
Qy 552 lyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArg----- 563
Db 1718 ACGGCAGCTTTGCGGCAAAACGGCAGCTACTTCCGCGCAGACCATCAAAAGACGCGCTTGCCA 1777
Qy 564 -----AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGly----- 578
Db 1778 ATCCGCAAAACCGGCACGACTCTGTGCGCGTCCGCGAAGCGGTCAACCGCGCTACATCA 1837
Qy 578 ----- 578

```

RESULT 5

```

US-10-085-959-54
; Sequence 54, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54
; LENGTH: 32159
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10795)..(10795)
; OTHER INFORMATION: Unsure
US-10-085-959-54

Alignment Scores:
Pred. No.: 7.4e-12 Length: 32159
Score: 220.50 Matches: 187
Percent Similarity: 34.05% Conservative: 97
Best Local Similarity: 22.42% Mismatches: 309

```



```

Db 30808 CAGTGAAGTTGATCAGCGTAACCGG-----GTGCCGTT 30840
Qy 658 gValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAl 678
Db 30841 CGTACCAAGTTATGGCCGGAGACAGCGGATGATGATGATGATGATGATGATGATGATG 30900
Qy 678 aAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgTh 698
Db 30901 A-----CTGATGCCCGGACTGGCGGTTAATCT 30927
Qy 698 rProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyG 718
Db 30928 GGTGGCCCGCATTTATTCGAT---GGCGAACACCGATGCGGACCGACCTATGCCAC 30984
Qy 718 u-----Ttp-----AsnTrpTyrValLysAlaAs 726
Db 30985 CCTGGACAGACGCTGGGCTGGCAGGCGACTGACGAGTGAACGATGACATTTCCGTCATGCGA 31044
Qy 726 pAsnLeuLeuAsnGln-----SerValTyrAlaHisSerSerPheLeuSerAspThrPr 744
Db 31045 TAACTGTTCACCGTGGTACCGTACCTATGCTGCTACATGAACGCGACGCGCGCTCGC 31104
Qy 744 oGlnMetGlyArgSerPheThrGlyGlyValAsnValLys 757
Db 31105 GCAGGTCAATATGGTGCACCGCTCGGTATCAATACGCGA 31144

RESULT 6
US-09-801-451A-1
; Sequence 1, Application US/09801451A
; Publication No. US20030096368A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; Beucher, Margaret
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule
; and Vaccine
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/801,451A
; FILING DATE: 08-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/418,964
; FILING DATE: 07-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-3-P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 318..2456
```

```

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-801-451A-1
Alignment Scores: 5,16e-12 Length: 2600
Pred. No.: 207.00 Matches: 196
Score: 34.66% Conservative: 100
Percent Similarity: 22.95% Mismatches: 308
Best Local Similarity: 5.13% Indels: 254
Query Match: 11 Gaps: 41
DB: 41

US-09-936-377-2 (1-758) x US-09-801-451A-1 (1-2600)
Qy 11 LeuSerIleLeuLeuIleAsnThrProLeu---LeuAlaGlnAlaHisGluThrGluGln 29
Db 339 CTCAGCTGCTCTCGCTCACACTTTCGCGCGCGCTTTCGCCACGCGGAGAAAATAATGCC 398
Qy 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGly 49
Db 399 AATGTCGATTTGGATACCGTTACCGTAAAGAGGACCGCCAA----- 440
Qy 50 LeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLys 69
Db 441 -----GGCAGCAAAATCCGTACCAACATCGTT-----ACGTTCAACAAAA 482
Qy 70 -----AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAla 85
Db 483 GACGAAAGACCGCAACCGATATCGCGAACTCTTAAAGAGAGAGCCCTCCATC----- 536
Qy 86 SerGlnTyrGlyGlyGlyAlaSerAlaProValIle-----ArgGlyGln 100
Db 537 ---GATTTTCGCGCGCGGCAACGCGACGTCCCAATTCCTGACGCTGCGCGGTATGGTCA 593
Qy 101 ThrGlyArgArgIleLysVal-----LeuAsnHisHis 111
Db 594 AACTCTGTCGACATCAAGGTGGACACACGCTATTCCGACAGCAAAATCTTTACCACCAA 653
Qy 112 GlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAla 131
Db 654 GGCAGA-----TTTATTGTCGATCCCGCT 677
Qy 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db 678 TTGTTAAAGTCGTTTCGCTACAGAAAGC-----GCGGTTTC 716
Qy 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu----- 169
Db 717 GCCTCTGCGGTATCGCGCGGACCAACGCGCGCATCATCGCAAAACCGTCGATCCCAA 776
Qy 170 -----AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSer----- 182
Db 777 GACCTGCTCAAAAGCTTGGATAAAACCTGGGGCGTGGCGCTCAACAGCGGCTTTGCCG 836
Qy 183 -----GlyAsnLeuGluLysLeu 188
Db 837 AACACGCGCGCAAGCTACGCGCAAGCTATTTCGAAAAGAGGCACTTCACGCTTTC 896
Qy 189 ThrSerGlyGlyIleAsn-----IleGlyLeuGlyLysAsnPhe-ValLeuHi 204
Db 897 TTCTCTTACAACCGCAACGATGAAAAGATTACGAAGCGCGCAAGGTTTCGCAATGAC 956
Qy 204 sThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeu 224
Db 957 AACG-----GCGGCAAAACCGTACCGTACGCGGCTCGCAACAGCGCAGC 1001
Qy 224 sArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTr 244
Db 1002 TACTCGCCAAAATCGGAACAACC-TTCGCGGACGCGGACCAACCGCATCTGTTGAGCCA 1060
Qy 244 pArgLysArgPheTyrArg-----ArgThrTyrSerAspArgAspGlnTyrGlyLe 262
Db 1061 TATGAAAGACCAACACCGCGGCGCATCCGCACTGTG-----CGTGAAGAGATTGCGCGT 1111
```

QY 262 uProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLe 282
 Db 1112 CAGCGAAAATAAT-----TACCGAT 1132
 QY 282 uileAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTy 302
 Db 1133 AACTATTAACGCGCAAGCCCATCTACCGGAA---ACCCTCAATCCAAACCAACTT 1189
 QY 302 rAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHis 332
 Db 1190 GCGGTACACCGGCAAGATTGGCTTTGTGAAATAACTGATGTCACCAACCTATGTGTT 1249
 QY 322 sAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLy 342
 Db 1250 -----GGAAAAAAGAACGCTATTCCGCCCATGACCAAGATAACGGCTACGAGCAATGTA 1305
 QY 342 sGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisH 362
 Db 1306 AAGGCCCAACCA----- 1318
 QY 362 sAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaAr 382
 Db 1319 -----TACCCGAATCGCCACTCGGAGTATGA 1344
 QY 382 gIleGluLeuArgHisGlnPro-IleGlyArgLeuLysGlySerTrpGlyValGlnTyrL 402
 Db 1345 ACTTCACTTCGACAGCGCTTCGCCGAACAAACCTGTTGAAATACGGCATCAACTACC 1404
 QY 402 euglyGln-----LysSerSerAlaLeu----- 409
 Db 1405 GCCATCAGGAATCAACACCGACGCTTTTGAATCGGAATTTGAATAAAGATAAAG 1464
 QY 410 --SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisT 429
 Db 1465 AAAAAAGCACTAATGAGAGAAAGAAAGAACCGTGAATAATGAAAAATTCGCAAGACCT 1524
 QY 429 yrSerPhePheGlyValGluGlnAlaAsnTrp----- 439
 Db 1525 ACCGCTGACCAACCGCAACCAACCGATACCGCGGCTATATGCAAGCCATTACAGAGA 1584
 QY 440 --AspAsnPheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgT 459
 Db 1585 TTGACGGCTTTACCTGACCGGGGCTGCTTACGACCGCTTCAAGGTGAAAACCCAC- 1643
 QY 459 yrAspLysAlaLeuIleAspArgGluAsnTyrTyrLys-----GlnProLeuProAspLeuG 478
 Db 1644 -----GACGG-CAAAACCGCTTTCAAGACGACGACCTCAACCCGAGTTT- 1684
 QY 478 lyAlaHisArgGlnThrAlaArgSerPheAlaLeuSer-GlyAsnTrpTyrPheThrPro 497
 Db 1685 -----CGCGGTGATTGCGCGCGCGGCAACACTGGAGCTTCAGCGCG 1727
 QY 498 GlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeu 517
 Db 1728 AGCCACAACCTAC-----GCGCGCGCGACGCGCGCTG-----TATGAGCTCTG 1772
 QY 518 TyrAlaHisGlyLysHis-----ValAlaThrAsnThrPheGluValGlyAsn 533
 Db 1773 CAACCCACGCGAAGCGCGCATCATCTCGATTGCGCGACGCG----- 1817
 QY 534 LysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAsp 553
 Db 1818 -----AAAGCCGACGCGCGCAATACCGAATCGCTTCAACTACACGACGCG 1868
 QY 554 ArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGln 573
 Db 1869 ACCTTTCCCGCAAA-----GGCAGCTAC---TTCGCGCAG 1901
 QY 574 ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuVal 593
 Db 1902 ACCATCAAGACGCGCTTGCCAAATCCGAAACCGCCACGACTCCGTCGCGCGCGGAA 1961
 QY 594 ArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGlu---GlyGluIleTyrPheLys 612

Db 1962 GCGGTACACCGCGCTACATCAAAAACCGGTTACGAATGGCGGCTCTCCGCGACC 2021
 QY 613 ProThrProArgTyrArgIleGlyValSer----- 622
 Db 2022 GCGCGCTGACCGCCCAAGAGTGGCGTAAGCACACAGCAAAACCGCGCTTTTACGATACGAC 2081
 QY 623 -----GlyAspTyrValArgGlyArgLeuLys 631
 Db 2082 AAAGACAAGCTGTGTGACGCGGAACCTCGAATTTGGCGCACAAGT-CGGCGCACTTGGAC 2140
 QY 632 AsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGln 651
 Db 2141 GGCTTCCTTCCTTACCGCTTCAAAAACCCGAATCTGGAATCGG-----CTGCGCGCG 2194
 QY 652 AspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHis-LeuLysAlaSe 671
 Db 2195 TCGTTATGTTCAAAAACCGCT-----GGTTTCGATATTGGCGCGCAGG 2236
 QY 671 rLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLe 691
 Db 2237 TCAAAAAGACCGC---GACGCGCAATTTGGAA-----AACGTTGTACGCCAA- 2279
 QY 691 uAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrAr 711
 Db 2280 -----GGTTTCGGTGTGAACGATGCTTCGCCCAACTCGAA 2314
 QY 711 g-----ArgAsnThrArgTyrGlyGluTyrAsnTrpTyrValLysAlaAspAsnLe 728
 Db 2315 ACCGTGGCGCAAGACGACG-----CTCAATGTTAACTTTTCGGTTAACACAGT 2362
 QY 728 uLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr---ProGlnMetGl 747
 Db 2363 GTTCGACAGATTCTACTATCCGACAGCAACGCTGACCAATAACCTGCGCGCGTGGG 2422
 QY 747 YArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 2423 ACGTGATGTACGCTGGCGGTGAACCTACAAAGTTC 2456

RESULT 7
 US-10-181-319-1
 ; Sequence 1, Application US/10181319
 ; Publication No. US20030135032A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lewis, Thomas A.
 ; APPLICANT: Paszczynski, Andrzej
 ; APPLICANT: Crawford, Ronald L.
 ; APPLICANT: Cortese, Marc S.
 ; APPLICANT: Sebat, Jonathan L.
 ; TITLE OF INVENTION: Compositions and Methods for Bioremediation
 ; FILE REFERENCE: IDAHL19440
 ; CURRENT APPLICATION NUMBER: US/10/181,319
 ; PRIOR FILING DATE: 2002-07-15
 ; PRIOR APPLICATION NUMBER: PCT/US01/02386
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: 60/177,251
 ; PRIOR FILING DATE: 2000-01-20
 ; NUMBER OF SEQ ID NOS: 35
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 2061
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas stutzeri
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2061)
 ; OTHER INFORMATION: ORF K
 US-10-181-319-1

Alignment Scores:
 Pred. No.: 3,11e-11 Length: 2061
 Score: 198.50 Matches: 161
 Percent Similarity: 36.08% Conservative: 106

Best Local Similarity:	21.76%	Mismatches:	283
Query Match:	4.92%	Indels:	192
DB:	12	Gaps:	36
US-09-936-377-2 (1-758) x US-10-181-319-1 (1-2061)			
Qy	42	SerArgProAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle	61
Db	179	TCACAGCCGGAATAATCGAGCGCG-CTGAAAGGGTGCCTCCACGCGTGGCGGTATC	237
Qy	62	SerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGly---Val	80
Db	238	GATGGCTGGACCGCCGAGCAGTCAGGCATCATAGCCTCAACAACTGGAAGACGCATT	297
Qy	81	ProGlyIleHisAlaSerGlnTyrGly---GlyGlyAlaSerAlaProValIleArgGly	99
Db	298	CCTGGTCTGTCATTCACAGCGTTCGGCAAGCAGGTATGAATTCACCGTCATCGCGGG	357
Qy	100	GlnThrGlyArgArgIleLysValLeuAsnHisGlyGlnThrGlyAspMetAlaAsp	119
Db	358	CTG-----ACGGCCAACTTCAACAGC	378
Qy	120	PheSerProAspHisAlaIleMetValAsp-----ThrAlaLeuSerGln-----	134
Db	379	TTCTCCAGTTCAACGTTGTTGCTGTCATGGCGTTCCTCCAGCTGACAGCCCGGATTC	438
Qy	135	-----GlnValGluLeuArgGlyProValThrLeuLeu	146
Db	439	GAGAGTGGCATGTGGATTCGATCGCATCGAGTTCATTCGGCGCGCAATCTACGCTG	498
Qy	147	TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro---Glu	165
Db	499	TATGGCGTAAATCCGAGCGCGGTG-----ATTGCCATCCACGCTGCCGATGAC	552
Qy	166	LysMetProGluAsnGlyValSerGlyGluLeuArg-----	179
Db	553	CGACCCCGAGAGCAGCGTGTCTGCCGAAGCGCGCAGCGGAACAACGCGTGTATCGCG	612
Qy	180	-----LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIle-----AsnIle	195
Db	613	TTTGGCTTCAGCCAGCCCTTGGTGAAGAGCGGTGTACGCGCAGCGTATCGGCAACTGG	672
Qy	196	GlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyr	215
Db	673	TCGAGCCAGGACGCTTATCGACACACCCAC---ACGGGGCACAGCGCGGACAT---	726
Qy	216	AlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsn	235
Db	727	-----CGTGACGAGAAACCTG---AACCTGGGGCTGCGTGGCGCCCG	768
Qy	236	GlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAsp	255
Db	769	GGGGCCCAACGATGTGTC-----ATGGCTATGGCATCAGGATGACGAGAT	819
Qy	256	ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlnTyrAspAspCysHisAlaAsp	275
Db	820	GGCGCTCCCTGTGGGCTCGCCGCGCGCCAAAGGAG---CGAGTCGCGTCCGGAACG	876
Qy	276	IleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeu	295
Db	877	CCGAGCTGGAAACCGTTCT-----	894
Qy	296	ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAsp	315
Db	895	-----GAGGGCCAGACCTTGCTTCATGTCCAGCATGAATTT	933
Qy	316	AspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAsnLeuArgAsnLysArgTyr	335
Db	934	GCCTCGCCCTCGCGTGTGATTCGGTAACGCGCTGGAACGAGTTCAAGGAC-----	984
Qy	336	GluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu	355
Db	985	-----AGGATTACAGAGGACACTGACTTATGCCAGCCGATGTCTG-----CACGTC	1032
Qy	356	AsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsn	375
Db	1033	GGCGCGAC-----CATCAC-----CTCGCG	1053
Qy	376	AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeu---Lys	394
Db	1054	ACACTCTCCAGGAGTTCGTTGGAG-----GGACAGCTCGGGAG	1095
Qy	395	GlySerTrp-----GlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThr	412
Db	1096	GCAGTTTGGCTGGCTGTGTC---TAGCGGATCGGACGACACATCTGCACAGTACC	1152
Qy	413	SerGluAlaLysGlnProMetLeuAspAsnLysValGlnHisTyrSerPhePhe	432
Db	1153	AGCAAGACCATGATGGG-----CTGCGACATTCGCGCGATCAGCAGCGATACC	1206
Qy	433	GlyValGluGlnAlaAsnTrpAsp-----AsnPheThrLeuGluGlyGly	447
Db	1207	GCTGCACCTGTCCACCACTGGAACGTCCTGCGCCGACTGGTCCATAGACCGCGGA	1266
Qy	448	ValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGlu	467
Db	1267	CGCGCTCGAGCGCAACGAGTGCAGTACGT---CCGCAAGGGGTACGAGCCATGAA	1323
Qy	468	AsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPhe	487
Db	1324	AAAGGCTGGACACACGCTTCCACCGAGCTCGCGTGCAACACACAG-----	1368
Qy	488	AlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis	507
Db	1369	-----ATAACGCGCAATCAC	1383
Qy	508	GlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn	527
Db	1384	CAATGCTATGAGTCCAGTCCGCGTGGCGTGGCGACTCGCGCTTCATGTGCTGGCGCG	1443
Qy	528	ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAlleGluLeuAla	547
Db	1444	ACGCTG-----GGTTATCTGCTTACGACACGAGAGAACTGGTCTGATGAA	1491
Qy	548	LeuGlyTyrGluGly-----AspArgTrpGlnTyrAsnLeuAlaLeuTyrArg	563
Db	1492	ACCGTCTCAGGCTGGCTTCTTGACAGCGATTCGCTATTGCTGGCGCGCTAC---	1548
Qy	564	AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSer	583
Db	1549	-----CTCATGGAC-----	1557
Qy	584	IleGluAspAspSerGluMetLysLeu-----ValArgTyrAsnGlnSerGly	599
Db	1558	ATCGATGACATGCAAGTTCATGCAAGATGCCACCGTCCGCGTGTATCATCACCGCGCT	1617
Qy	600	AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIle	619
Db	1618	GCCACGCGCATCCAAAGTCTCGAGCTGGATGGACTA-TCTCTCTGGTGGCGGCTG	1676
Qy	620	Gly-ValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyAr	639
Db	1677	GCAGTCAAGGGCGGCTGGCCCTGGAAACACACCGCTTCGATCATCTTCGCGATGGCA	1736
Qy	639	GluAspAlaTyrGlyAsnArgProPheIleAlaGlnAsp-----	652
Db	1737	GGCGGACTATGACGGCAACCAAGCCGTTTCGCGCGGATCTCCACCGCCACCTCGGCAT	1796
Qy	653	-AspGlnAspArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLe	672
Db	1797	CCGCTAGCAGCCCGCGAA-----GGCTGTATGCAACAGCCGCT	1838
Qy	672	uThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAl	692
Db	1839	GACCGCGCAGCAGCAAGTCTACCTCGATGCGGCGCAACGGGTATGAACGCAAC-----	1890


```
Db 2126 CTTTCCGCGACAGTTTCCTCTCCTACTACTGACATATGG----- 2088
Qy 349 GluAlaLeuArgValHisLeuAsnArg-AsnAspTyrHisHisAspGluLysAlaGlyAs 368
Db 2087 -----CGGTGAGTATTATCGTCAGGAACAACATCCGGCGG 2052
Qy 368 pAlaValGluAsnPhe-----PheAsnAsnGlnThrGlnAsnAlaAr 382
Db 2051 CGCGACGACGGGCTCCGCAAGCAAAATCGATTTCGTCCGGTGGCTACAGGATGA 1992
Qy 382 gIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLe 402
Db 1991 GATCACCTTACGGAGATCTCGCGATT----- 1967
Qy 402 uGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLe 422
Db 1967 ----- 1967
Qy 422 uAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAspAsnPh 442
Db 1967 ----- 1967
Qy 442 eThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAl 462
Db 1966 -ACCTGCTTGGCGGACCGCTATGACAGTTATCGCGGTAGC----- 1925
Qy 462 aLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGl 482
Db 1924 -----AGTGACGGTTACAAA-----GATGTTGATCCGCAAAATG 1890
Qy 482 nThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSe 502
Db 1889 GTCATCTCGTGGC-----GGGATGACTATCAATCCGCACTAACTGGGTGAT 1845
Qy 502 rLeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHi 520
Db 1844 GTTATTGGTCTCATATCCCGAGCATTCGCGCGCCGACGATGGCGGAATGTATAACGA 1785
Qy 520 sGlyLysHisValAlaThrAsnThrPheGlu-----ValGlyAsnLysHisLe 536
Db 1784 TTCTAAGCACTTCTCGATTGGTTCGCTTCTATACCACTATTGGGTGCAAAACCGAATT 1725
Qy 536 uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly----- 549
Db 1724 ACCTCCGGAACCTAACGAACCTCAGGAGTACGGTTCGTTGGGCTCGCTTTTGATGACCTGAT 1665
Qy 550 -TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569
Db 1664 GTTGTCCAATGATGCTCTGGAATTTAAAGCCAGCTACTTTGATACCAACGGAAGGATTA 1605
Qy 569 rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGl 589
Db 1604 CATC-----TCCACGACCGTCTGATTCGCGGC 1578
Qy 589 uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla-----GluGl 607
Db 1577 GCGCAGCACTATCGTATTAACGTCGCCGACGCAAAATCTGGGCTG-GGATGTGATGA 1519
Qy 607 yGluIleTyr-----PheLysProThrProArgTyrAr 618
Db 1518 CGAAATATACCACTGATCTCTTTAGCCTTGATCGGCTATAACCGTACC CGCGCAAG 1459
Qy 618 gIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGl 638
Db 1458 A-----CACGATACCGCGAATATACATCTCCACATTAACCCCGGATACTGTTA 1411
Qy 638 yArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProAr 658
Db 1410 CCAGCACTCTGAATATCCGATCGCTACA-GTGGCTTCTCTGTGGGTGGTGTGTACG 1352
Qy 658 gValProAlaAArgLeuGlyPheHisLeuLysAlaSerLeuThr 673
Db 1351 TTTGCCGATCGCTCAA-----CACATATCAGCAGCAGTTACA 1315
```

RESULT 9

```
US-10-238-075-731
; Sequence 731, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isola
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 731
; LENGTH: 1980
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-731
```

```
Alignment Scores:
Pred. No.: 8,84e-10 Length: 1980
Score: 185.00 Matches: 166
Percent Similarity: 31.97% Conservative: 92
Best Local Similarity: 20.57% Mismatches: 265
Query Match: 4.58% Indels: 286
DB: 12 Gaps: 35
```

US-09-936-377-2 (1-758) x US-10-238-075-731 (1-1980)

```
Qy 4 ThrThrLysLysProIleValLeuSerIleLeuLeuIleAsnThr---ProLeuLeuAla 22
Db 19 ACCTCGTGGTGTGAGTTGTTGGCTTGTGGCTGTTTCCACCTTCCACACGTTTGCT 78
Qy 23 GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValGlyLysSer 42
Db 79 TTTGCTACTGAAACC-----ATGACCGTTACGGCAACGGGAATGCA 120
Qy 43 Arg-----ProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSer 57
Db 121 CGTAGTTCCTCGAAGCGCTATGATGGTCAGC---GTTATCGACACTTCGCTCTCTGAA 177
Qy 58 AspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeu 77
Db 178 AAT-----CAACTGCTACTTCAGCCACTGATTTCTG 210
Qy 78 AspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal--- 96
Db 211 CGTCATGTTCTCGAATTAATCTTTGATGGTACCGGACGAACCAACGGTCAGGATGTAAT 270
Qy 97 IleArgGlyGlnThrGlyArgArgIleLysValLeu-----AsnHisHisGly 112
Db 271 ATCGTGGCTATGATCATCGCGCGTCTGTTCTTGTGATGTTGTCGCCAGGACG 330
Qy 113 GluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu 132
Db 331 GATACCGGACACCTG-----AATGGCACTTTTCTCGATCCGGCGCTG 372
Qy 133 SerGlnGlnValGluIleLeuArgGlyProValThrIleuLysTyrSerSerGlyAsnVal 152
Db 373 ATCAACGCGTGTGAGATTGTTCCGGACCTTCAGCATTTACTGTATGGCAGTGGCGCTG 432
Qy 153 AlaGlyLeuVal-----AspValAlaAspGlyLysIleProGluLysMetProGluAsn 170
Db 433 GGTGGAGTATCTCCATACGATCGTCGATGCAAAA-----GATTATTGCGAGAA 483
Qy 171 GlyValSerGlyGluLeuGlyLeuArgLeu---SerSerGlyAsnLeuGluLysLeuThr 189
Db 484 GGACAAAGCAGT-----GGTTTCTGTTCTTTGGTACTTGGCGGC----- 522
Qy 190 SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu--- 208
```


Pred. No.:	8.63e-09	Length:	9054
Score:	185.00	Matches:	166
Percent Similarity:	31.97%	Conservative:	92
Best Local Similarity:	20.57%	Mismatches:	265
Query Match:	4.58%	Indels:	286
DB:	12	Gaps:	35

US-09-936-377-2 (1-758) x US-10-085-959-172 (1-9054)

Qy

4

ThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThr---ProLeuLeuAla 22

Db

3097

ACCTCGTTGGTTGAGTTTGTTGGCTTGGCTTTCTGCCACCTTGGCCACGTTTGCT 3038

Qy

23

GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSer 42

Db

3037

TTTGCTACTGAAACC-----ATGACCGTTACGCAACGCGGAATGCA 2996

Qy

43

Arg-----ProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSer 57

Db

2995

CGTAGTTCCCTCGAAGCCCTATCATGTGTACG---GTTATCGACACTTCGCTCCCTGAA 2939

Qy

58

AspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeu 77

Db

2938

AAT-----CAAACCTGCTACTTCAGCCCACTGATTTCGTG 2906

Qy

78

AspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal--- 96

Db

2905

CGTCATGTTCTGGAATTACTTCTTGATGTGACCGGACCAACCAACGGTCAGGATGTAAAT 2846

Qy

97

IleArgGlyGlnThrGlyArgArgIleLysValLeu-----AsnHisHisGly 112

Db

2845

ATGGTGGCTATGATCATCGCGGCTGCTGTTCTTGTGATGGTGTGCCAGGGAACG 2786

Qy

113

GluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu 132

Db

2785

GATACCGGACACCTG-----AATGGCACTTTCTCGATCCGCGCGCTG 2744

Qy

133

SerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnVal 152

Db

2743

ATCAAGCGTGTGAGATTGTTTCGCGACCTTCACATTTACTGTATGCACTGGCGCGCTG 2684

Qy

153

AlaGlyLeuVal-----AspValAlaAspGlyLysIleProGluLysMetProGluAsn 170

Db

2683

GGTGGATGATCTCCTACGATACGGTCATGCATAAA-----GATTATTGCAGGAA 2633

Qy

171

GlyValSerGlyGluLeuGlyLeuArgLeu-----SerSerGlyAsnLeuGluLysLeuThr 189

Db

2632

GGACAAAGCAGT-----GGTTTTCGTGTCTTGTGTACTGGTGGCGC----- 2594

Qy

190

SerGlyGlyIleAsnIleGlyLeuGlyAsnPheValLeuHisThrGluGlyLeu--- 208

Db

2593

ACGGGGACCATAGCCTCGGATTAGGCGGACGCGGTTTGGGCGAACTGAAATCTGGAT 2534

Qy

209

-----TyrArgLysSerGlyAspTyrAla 216

Db

2533

GGTATTGTGGCTCGTCAGTCGGATCGGGTGATTATGACCAACCAATGTGTAACC 2474

Qy

217

ValProArg----- 219

Db

2473

GCGCCGAATGACGAGTCCATTAAATACATGTCGGCGAAAGGGACCTGGCAAAATTGATTCA 2414

Qy

220

-----TyrArgAsnLeuLysArgLeuProAspSer 229

Db

2413

GCCAGTCTCTGACGGTTTAGTCGGTTACTTACAAACCAACGCGCGCTGAACCAAAAAAT 2354

Qy

230

ProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPheTyr 249

Db

2353

CCGACAGCCTTGAGCTTCTGTAGCAGCAACCCGATGTGTGATCTGTCATCAACAATTCAA 2294

Qy

250

ArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyr 269

Db

2293

CGC-----GATGCGAGCTTTCTTATAAACTCGCCCGCAGGGGTACGACTGG 2246

Qy

270

AspAspCysHisAlaAspIleIleThrGlnLysSerLeuIleAsnLysArgTyrLeuGln 289

Db

2245

TTAAATTCAGATCAAAAAATTTTACTGTCGGAAGTCCGTATTATATGCGCAA----- 2195

Qy

290

LeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCys 309

Db

2194

-----AACACGGGGAGTTTCAGGC 2177

Qy

310

GlyPheHisAspAspAlaHisAlaHisAlaHisAsnGlyLysProTyrIleAsp 329

Db

2176

GAGTATCGTGAACA-GATAACAAAGAGCAAGGCT-----CGAGAACCGTTCCACTCT 2124

Qy

330

Leu---ArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPhe 348

Db

2123

ATTTCGCGACAGTTTCGCTTCTCACTTACTGACATATGG----- 2085

Qy

349

GluAlaLeuArgValHisLeuAsnArg-AsnAspTyrHisHisAspGluLysAlaGlyAs 368

Db

2084

-----CGGTGAGTATTATTCGTACAGAACCAATCCGGGTGG 2049

Qy

368

pAlaValGluAsnPhe-----PheAsnAsnGlnThrGlnAsnAlaAr 382

Db

2048

CGCGACGACGGGCTTCCCGCAACAAAAATCGATTTTACTGCTCTGGTGGCTACAGATGA 1989

Qy

382

gIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLe 402

Db

1988

GATCACCTTACGCGATCTGCCGATT----- 1964

Qy

402

uGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLe 422

Db

1964

----- 1964

Qy

422

uAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAspAsnPh 442

Db

1964

----- 1964

Qy

442

eThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAl 462

Db

1963

-ACCTCTGTTGGCGAACCCTGATGACATTTATCGCGGTAGC----- 1922

Qy

462

aLeuIleAspArgGluAsnTyrTrpLysGlnProLeuProAspLeuGlyAlaHisArgGl 482

Db

1921

-----AGCGACGGGTACAAA-----GATGTTGATCCGCAAAATG 1887

Qy

482

nThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSe 502

Db

1886

GTCACTCTGTGG-----GGGATGACTATCAACCCACCACTGGGTGAT 1842

Qy

502

rLeuThrAlaSerHisGlnGlu-----ArgLeuProSerThrGlnGluLeuTyrAlaHi 520

Db

1841

GTTATTGGCTCATATGCTCAGGCATTCGCGCCCGACGATGGCGGAAATGATAACGA 1782

Qy

520

sgLysLysHisValAlaThrAsnThrPheGlu-----ValGlyAsnLysHisLe 536

Db

1781

TTCTAAACACTTCTCGATTGGTCGCTTCTATACCAACTATTGGTGGTCCAAACCGCAACT 1722

Qy

536

uAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly----- 549

Db

1721

ACGTCCGGAACATAAGAACTCAGAGTAGCTGGTTTTGGCTCGTTTTGATGACCTGAT 1662

Qy

550

-TyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTy 569

Db

1661

GTTGTCCAATGATGCTCTCGAATTTAAAGCCAGCTACTTTGATACCAAAACGGAAGATTA 1602

Qy

569

rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGl 589

Db

1601

TATC-----TCCACGACCGTGGATTTCGCGGC 1575

Qy

589

uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla-----GluGl 607

Db

1574

GGCGCAACATATGTCGTATTAACGTCCCGCAACCAAAATCTGGGGCTG-GGATGTGATGA 15

```

Db      1515  CAAAATATACCACTGATCTGTTTAGCCTTGATGGCCTATAACCGTACCCGCGCAAG 1456
Qy      618   gileGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGl 638
Db      1455  ACACCGATACCGGGGAATATAT-----CTCCAGCATTAACCCGG 1417
Qy      638   yArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaPro-A 658
Db      1416  ATACCGTTACCACTAC-----CCTGA 1396
Qy      658   rgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspA 678
Db      1395  ATATTCGGATCGCTACAGCGGCTTC----- 1370
Qy      678   laAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgT 698
Db      1369  --TCTGTTGGTGGTCCGTACGTTGCCGATCCTCA----- 1334
Qy      698   hrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyG 718
Db      1333  -----ACATATATCAGCAGCTACAGCAACAAACACCTGGCTATGGTG 1291
Qy      718   luTrpAsnTrpTyrVal 723
Db      1290  TGAATGATTTCTACGTC 1274

RESULT 11
US-10-238-075-507/c
; Sequence 507, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 507
; LENGTH: 12264
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-507

Alignment Scores:
Pred. No.: 1.75e-08 Length: 12264
Score: 184.00 Matches: 134
Percent Similarity: 37.06% Conservative: 108
Best Local Similarity: 20.52% Mismatches: 236
Query Match: 4.56% Indels: 176
DB: 12 Gaps: 35

US-09-936-377-2 (1-758) x US-10-238-075-507 (1-12264)
Qy      1   MetAlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db      10199  AIGTATGAAATGTAATCAGAACTGATTTGTATTAATAATACTT-----CCGGTG 10146
Qy      21   LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
Db      10145  GGATTACAGCGCAGCCAGTCATTCTTCTATGGTTAAAGATACAATCACCATTGTCTCGG 10086
Qy      41   LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db      10085  -----ACAGGAAATCAGAACACCGGTATTTGAAACGCCGTCGATGGTCAGTGTC 10038
Qy      61   IleSerGlyAspThr---LeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGly 79
Db      10037  GTCGAATGACACACCGTGGAGTCAAGATCGGTCATCATCGCCGCGCATGCTGAAAGGT 9978

```

```
Qy 411 aThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPh 431
Db 9135 TGCCACAC-----CTATATCCGGAAGAAACATGCTTACATCGTT 9094
Qy 431 ePheGlyValGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValG1 451
Db 9093 GTATTTCAGATGAATGACAATGAAAGCTACCGGTTAACATTTATCGTC----- 9042
Qy 451 uLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLy 471
Db 9041 -----GGTTCGCGCTATGACCG-----TACAA 9019
Qy 471 sGlnProLeuPro-----AspLeuGlyAlaHisArgGlnThrAlaArgSerPheAl 488
Db 9018 GAGCTTCAATCCCGTCGCGGAGAACTGAAAGCCGACGCTGTCGCCAAGGGCG-----GC 8962
Qy 488 aLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeu-----ThrAlaSe 506
Db 8961 GATTTC-----GTCTCACCACAGACTGGCTGATGATGATGACGGCTCCATATC 8914
Qy 506 rHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaTh 526
Db 8913 CTCTGCATTCGAGCGCCCAATGCGCAGAAATGTACAGGATGATGATCATTTT----- 8859
Qy 526 rAsnThrPheGluValGlyAsn-----LysHisLeuAsnLysGluArgSe 541
Db 8958 -----TACCGCAAGGTAACCAATTACTGGGTCTTCAACCTTAATCGAAACCCAGA 8806
Qy 541 rAsnAsnIleGluLeuAlaLeuGly-----TyrGluG1 552
Db 8805 AAATAACATCACCGTGAGATTGGCGCAGGATTCAACTGGATGGCTGCTTACAGACAA 8746
Qy 552 yAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAl 572
Db 8745 TGACCGGTCGAGTTAAAGCGGATATTTCGGAACGGATGCCAGAACTATATTGCCAC 8696
Qy 572 aGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLe 592
Db 8685 ACGCGTG-----GATATGAAACGGATCGGTC 8659
Qy 592 uValArgTyrAsnGlnSerGlyAlaAspPheTyrGly 604
Db 8658 TTATTCTTAAATGATATCCCGGCGCGTATCGGGGA 8622
RESULT 12
US-10-085-959-11
; Sequence 11, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 23654
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION: Unsure
US-10-085-959-11
```

```
Alignment Scores:
Pred. No.: 4.69e-08 Length: 23654
Score: 184.00 Matches: 134
Percent Similarity: 37.06% Conservative: 108
Best Local Similarity: 20.52% Mismatches: 236
Query Match: 4.56% Indels: 176
DB: 12 Gaps: 35

US-09-936-377-2 (1-758) x US-10-085-959-11 (1-23654)

Qy 1 MetaLalGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuLeuLeuLeu 20
Db 2065 ATGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2118
Qy 21 LeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGly 40
Db 2119 GGATTACAGGACGACGACGATCTCTCTATGTTAAAGATACATCAATCACCATTGTGCGG 2178
Qy 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db 2179 -----ACAGGAAATCAGAACACGCTATTGAAACGCGCTCGATGTCAGTGTGTC 2225
Qy 61 IleSerGlyAspThr-----LeuArgGlnLysAlaValAlaAsnLeuGlyAspAlaLeuAspGly 79
Db 2227 GTACAGATGACACACCGTGGAGTCAGATCGGTTACATCGCGCGGATCTCTGAAGGT 2286
Qy 80 ValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGly 99
Db 2287 GTTCCGCGTCTC-----AGCCAGACTGGTCAGACGGACC-----AATGGG 2328
Qy 100 GlnThrGlyArgGlnLysValLeuAsnHisGlyGluThrGly----- 115
Db 2329 CAGACCTTTAAT-----TTACGCGGCTATGACAAAGCGGGTACTTGTTCCTT 2376
Qy 116 -----AspMetAlaAspPheSerProAspHisAlaIleMetValAspThr 130
Db 2377 GTTCACGCGGTTCCCAACTCAGTGACATGCAAAAAGCAGTGGCAGCTTATTCGATCCG 2436
Qy 131 AlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuLysSerSerGly 150
Db 2437 GCACCTGCAACGATATCGAAGTTGTCGCGGGGCCAAACTCCAGCTGTACGCGAGTGGC 2496
Qy 151 AsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsn 170
Db 2497 GGGCTGGGAGGTGATGAGACTTCAGAACTGCCGATGCAGCAGATTTTCTTCCCGCCGGA 2556
Qy 171 GlyValSerGly-----GluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThr 189
Db 2557 GAGACAAACGGTTTAACTCTGTGGGGAATATCCCGAGTGGT-----GACCACAGCACA 2610
Qy 190 SerGlyGlyIleAsnIleGlyLeuGlyLys-----AsnPhe 201
Db 2611 GGCTCGGGGCTCACC---TGCTTTGGTAAACTGGAATAACAGATGCGCTCTTCTCTGTC 2667
Qy 202 ValLeuHisThrGluGly-----LeuTyrArgLysSerGlyAspTyrAlaValProArgTyr 220
Db 2668 ATTATGCGTAAAGAGGATATATCTATCAAGTGTGTCAGCAGCA---CCTAAC--- 2721
Qy 221 ArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGly----- 236
Db 2722 -----AAGGNAACCTGACGCC-----CTGTTTGGAAAGGCTGTGTCGGTATA 2766
Qy 237 -----GlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgTyrSer 254
Db 2767 ACAGACAGTAACAAAGCA-----GGTGCCAGCTTGGTCTCTACCGGAATAACACCAT 2820
Qy 255 AspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAla 274
Db 2821 GAA-----CCGGCAATTCCACTCAG-----ACACATGGT 2850
Qy 275 AspIleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeu 294
Db 2851 GAC-----ACGCGCTGCGTGACAGAAACACAGTACAA----- 2883
```



```

QY 611 -----
Db 4428 TGCTTATAACCGAGTAAAGTTAAAGATCAAAAAATCAATGCTGGTTCCTCCGTAAG 4487
QY 612 -----LysProThrProArgTyrArgIleGlyValSerGlyAs 624
Db 4488 CAGTTATTATTGATGCCATTCAGCCAGC---CGTATATCATTTGGTTTAGGCTATGA 4544
QY 624 p-----TyrValArgGlyArgLeuLysAsnLe 633
Db 4545 TCATCCAAGTAATACITGGGGAATTAATACAATGTTTACTCAATCAAAAGCAAAATCTCA 4604
QY 633 uProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgPropheIleAlaGlnAspAs 653
Db 4605 AAATGAATTCTAGGAAACGTT---GCATTGGGTAACAAT-----TCAAGGGATGT 4652
QY 653 pGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuTh 673
Db 4653 AAATCAACAGAAACTTACTCGGCA-----TGCAATATC----- 4689
QY 673 rAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaAr 693
Db 4690 -----TTAGATGATATCGGGTTATTACATGCGCAATAAA----- 4722
QY 693 gTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgAs 713
Db 4723 -----AAATATTGCTTCGATTAGGGATATATTAATTATTCAA 4760
QY 713 nThrArgTyrGlyGluTrpAsnTrpTyrValLys---AlaAspAsnLeuLeuAsn----- 730
Db 4761 CTATCGTATGTTACTTGGGAAGCGGTGCGTCAAAACAGACACAAGGTGCGGTCAATCAACA 4820
QY 731 -GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPh 750
Db 4821 TCAAAATGTTGGTAGCTATATCTCGTACGACGATCA-----GGACGAAACTA 4868
QY 750 eThrGlyValAsnValLysPhe 758
Db 4869 TACCTTAACATTGAAATGAATTC 4893

RESULT 15
US-09-815-242-4030
; Sequence 4030, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4030
; LENGTH: 2442
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4030

Alignment Scores:
Pred. No.: 6,25e-09 Length: 2442
Score: 178.50 Matches: 180
Percent Similarity: 35.73% Conservative: 103
Best Local Similarity: 22.73% Mismatches: 279
Query Match: 4.42% Indels: 233
DB: 9 Gaps: 43

US-09-936-377-2 (1-758) x US-09-815-242-4030 (1-2442)

QY 41 LysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIle 60
Db 374 AAATCAACCGCTCGCGC-ACCAGCAGCGACCTGGTCTCGCCACGCGCCAGTCCACGCTC 432
QY 61 IleSerGly-----AspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeu 77
Db 433 ATCGACATGCCCACTGGAGGAGCTGGCGCAGGGCTCCGACAGCTGGCCACCGTCTG 492
QY 78 ---AspGlyValProGlyIleHisAlaSer-----GlnTyrGlyGly 91
Db 493 GCCAAGCGGTGCGGGATGTCGATTCAGCGCCACCATCACCGAATAC----- 543
QY 92 AlaSerAlaProValIleArgGlyGlnThr-----GlyArgGlyLeuValLeuAsn 109
Db 544 -----GGCCAGACCTCGGGGGCGCAGCATGCTGGTGGTATGGTC 582
QY 110 HisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAsp 129
Db 583 GACGGCGTGGCTGAACACCAACCGCGACTCTCGCGCAACCTGGCC---AACATCGAC 639
QY 130 ThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSer 149
Db 640 CCGCGCTGATCGAGCGGATCGAGTCACTCGTGGCAGC---AGCGCCATCTACGGCAGC 696
QY 150 GlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu 169
Db 697 GCGCCACCGCGGGATCATCTCCATCAC---ACCGTCCGCGCC 738
QY 170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsn---LeuGluLysLeu 188
Db 739 GCGCGC---GAGAACCGCGGGAAACCGCGCTCAGCGCCACCTCGCGCTGACCCGCTG 795
QY 189 ThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu 208
Db 796 GCGAGCGAT-----GGCCTCGGGCGCCAGTTC----- 822
QY 209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAsp 228
Db 823 -----CAGCAATA-CTTCGCGCGCTCCCTGGG 848
QY 229 SerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe 248
Db 849 GCGCGCTCGACTATTCTGTTGACTTCGACCTCGGCACCCGCCA-----CGTGGCGCTTC 896
QY 249 TyrArgArgThrTyrSerAspArgArg-----AspGlnTyrGlyLeuProAlaHis-Se 266
Db 897 CTACGACGCCCATGGCGACCGCATCGCCCGGAAACCCAGCCAGCGGACACTGTTTCGACTC 956
QY 266 rHisGluTyr-----AspAspCysHisAlaAspI1 276
Db 957 GAACGCTCAACATCGCGCGCAAGCTCGCCCTGGCATCGACGAG----- 1002
QY 276 eIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuTh 296
Db 1003 -----AACGAGCGGTGACGCTCGCCCTCAGCCACTACGACGC 1040

```

QY 296 rGluGluAspValAspTyrAspAsn-----ProGlyLeuSe 308
 Db : : : : :
 1041 CGCCAGACACCGGACCTAGCCACCGAGGTCGCCAGCGTCCGCCCGCGC----- 1095
 QY 308 rCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTyrP1 328
 Db : : : : :
 1096 -----TCGGTCCCGCCCAACGGGATCAAAAGGCTGGAG---CT 1130
 QY 328 eAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPh 348
 Db : : : : :
 1131 GGACGAGCAGAACCGCATCCGCAACACCTCGCGAACCCTCGAG-----TACGAGAACCT 1184
 QY 348 eGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAsp----- 363
 Db : : : : :
 1185 CGACATCTCCGCGACCGGCTCTCCGCGAGCTCTACTACCGCGACTATTTACCCGCGTT 1244
 QY 364 -----GluLysAlaGlyAspAlaValGluAsnPhePh 374
 Db : : : : :
 1245 CACTCGTTTCGACCGCGCGGTCTCCACCGCGCGGCAATGTCCGACGATCATGCA 1304
 QY 374 eAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLy 394
 Db : : : : :
 1305 GAACAGCGAAGTGTTCGGCAGCGCGCTGACCTCGGC---ACCGCGCTGGCGGAAGCGG 1361
 QY 394 sGlySer-----TyrGlyValGlnTyrLeuGlyGlnLysSer----- 406
 Db : : : : :
 1362 CAATACCGAAGTGTCTGGGGGGGCGACTACAAACGGAACGACGAGCATCGCGCTCGA 1421
 QY 407 -----SerAlaLeuSerAlaThrSerGluAlaVal-----Iy 417
 Db : : : : :
 1422 CGTGTTCGACCGCGCGCTACGAGCCAGCGCGGGCTGGTCTTCGACAGATCGGCA 1481
 QY 417 sGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAl 437
 Db : : : : :
 1482 GCTCACTACATGCGCGCGCTACGACCGCGCGCGCGCGGTTCGCCAGTTTCGACGA 1541
 QY 437 aAsnTyrAsp-----AsnPheThrLeuGlyGlyValArgValGluLysGlnLysAlaSe 456
 Db : : : : :
 1542 TCGCTTCGACGACACTGTGTGATGACGCGCGCGCTGCGCTACGATATCTCCACGCGCA 1601
 QY 456 r-----IleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTy 470
 Db : : : : :
 1602 ATTCGACGACTTCATCCGCTCTCCGAGTCCAAAGCGCG-CTCGCGGTGACCGTCAAGG 1660
 QY 470 rLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSe 490
 Db : : : : :
 1661 GCGGCGACTCGACTACGACGCGGTGTCTCGAAACCTCGCGCATCGTCTATTTCGCC----- 1715
 QY 490 rGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluAr 510
 Db : : : : :
 1716 -GGTGGCTGG-----CCAGGAATATATGCCCTCTT-----CAGCCAGG 1753
 QY 510 gLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThr-AsnThr----- 528
 Db : : : : :
 1754 GCTTCCA-----GTTGCCGATGTTCGCATCCAGTTCGCGCAACGCCCGCC 1798
 QY 529 -----PheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAnilleGluLeuA 547
 Db : : : : :
 1799 GTGGCTTCGATATCGCTCTCTCGAACCCTCGAACCCGCGTGAAGCAACCACTACGAACTCG 1858
 QY 547 la---LeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyr-----ArgA 564
 Db : : : : :
 1859 GTGGCGTGGCCCATCGCGGCAACACCTCGGCGAGCTCGCGCTGTCTACACCACT 1918
 QY 564 sNArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerI 584
 Db : : : : :
 1919 CCAAGCTGGCGAC-----GTGCAGAGCTTCAACCAACGCG----- 1953
 QY 584 leGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrG 604
 Db : : : : :
 1954 -----CTGATCCTCACCAGCCGCAACCAAGGAGCGC-----ATCTACG 1987
 QY 604 lYalaGluGlyGluIle---TyrPheLysProThrProArgTyrArgIleGlyValSerG 623

Db 1988 GGTTCAGGCCAGCCGCGACTCGCTCAGCGACGAGAGTCTGGGCGCGGGCGGCGCG 2047
 QY 623 lyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAla 643
 Db : : : : :
 2048 CCACCTGGATCGCGGACCGGAGAG-----CCGGACGCGCAAGGAC----- 2088
 QY 643 yrGlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAla 663
 Db : : : : :
 2089 -----TGGCAGGACATGACCGGCTACCGGCTACCGGCTGCGCGCGCTGA 2125
 QY 663 rGLeuGlyPheHis-----LeuLysAla 671
 Db : : : : :
 2126 AGCTGACCGCCTACTCTGCAATACAAGCCGAGTGCAGACTGGAACAACCTCTGCGAGCCA 2185
 QY 671 erLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysL 691
 Db : : : : :
 2186 CCTTCTTCGACTCCCAAGGACTACCGTCTCGACGGCTGGAAAGCTTCGCGCGCGCAGG 2245
 QY 691 euAlaArgTyrGluThr-----ArgThrProGlyHisHisMetL 704
 Db : : : : :
 2246 TCAGCACCTTACACCGCGTCCGACCTGGTTCAGCCAGTACCGCATCACCCCGCAGCAGT 2305
 QY 704 euAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAsnTyrValL 724
 Db : : : : :
 2306 TGAGCTGGGC----- 2316
 QY 724 ysAlaAspAsnLeuAsnGlnSerValTyr 734
 Db : : : : :
 2317 --ATCCAGAACCTGTTCAACCGCGACTACTATC 2346
 RESULT 16
 US-10-238-075-262
 ; Sequence 262, Application US/10238075
 ; Publication No. US20030148324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238,075
 ; CURRENT FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 0003145
 ; PRIOR FILING DATE: 2000-03-10
 ; NUMBER OF SEQ ID NOS: 1576
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 262
 ; LENGTH: 2139
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 ; US-10-238-075-262
 Alignment Scores:
 Pred. No.: 6 59e-09 Length: 2139
 Score: 177.50 Matches: 167
 Percent Similarity: 31.4% Conservative: 103
 Best Local Similarity: 19.4% Mismatches: 283
 Query Match: 4.40% Indels: 305
 DB: 12 Gaps: 41
 US-09-936-377-2 (1-758) x US-10-238-075-262 (1-2139)
 QY 6 LeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
 Db : : : : :
 31 CTCAAAGGTCGGCGCTCTTTTCACTGCTTTTGGGACCGCATGATT-----CAT 81
 QY 26 GluThrGluGlnSerValGlyLeuGlu-----ThrValThrValValGlyLys 41
 Db : : : : :
 82 GCACAGACTCTGTAAACCAAGATGGCGAAACAATCACTGTTACAGAGATGCAAT 141
 QY 42 SerArgProArgAlaThrSerGlyLeu-----LeuHisThrSerThrAla----- 56
 Db : : : : :
 142 ACCGCAACTGAGGCAACCGATGTTTATCACTCTGAGCACTCCACGCGCAGCATTAACC 201

```
QY 57 -----SerAspLysIleIleSer 62
Db 202 GATATGCGGATGCTGGATATCCCGCAGGTGTCATACGGTTACCGATCAGGTTCTGGAA 261
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 262 -----AACCAGAAATGCACCAACGCTGATGAGCGCTTTTATAACGTCAGTAAC 309
QY 83 ---IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThr 101
Db 310 GTGTACAGACCAATACATAGGCGGAGCTAGATGCTTTTGACCGCGTGGGTTTGGC 369
QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGlnThrGlyAspMetAlaAspPheSer 121
Db 370 GCAAAACCGG----- 378
QY 122 ProAspHisAlaIleMetValAsp-----ThrAla 131
Db 379 ---GATGGCTCCATCATCAACCAACGGTCTGCGAACCGTACTTCTCGTAGTTTCAACGCC 435
QY 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db 436 GCACAGAGCGTGGGAAGTCTTAAGGCCCGCGCTCCACGCTGTATGGCAATTCTCGAT 495
QY 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly 171
Db 496 CCTGGCGGACTGATTAACTGCTGACC---AAGCGCCGGAAGAAACATTCATCGTGTGCG 552
QY 172 ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGly 191
Db 553 GTTTTCAGCC-----ACCTCTCC 570
QY 192 GlyIleAsnIleGlyLeuGlyLys-----AsnPheValLeuHisThrGluGly 207
Db 571 AGTTTGTGGCGGCACTGGCACTTGATATACAGGTCCCATTTGAAGGCACTCAGCTG 630
QY 208 LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro 227
Db 631 GCGTATCGCCCTACCGGGGAAGTG----- 654
QY 228 AspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg 247
Db 655 -----CAGATGAA 663
QY 248 PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHis 267
Db 664 GATTACTGGCGAAACTTCGGTAAAGAGCGCAGTACATTTATGCCCCGTCA----- 714
QY 268 GluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyr 287
Db 715 -----CTCACCTGGTTGGTGAT-----AATGCCAACAGTA 744
QY 288 LeuGlnLeuTyrProHis----- 293
Db 745 ACCATGCTCTATTCCCATCGGACATATAAACTCCATTGATCGTGGACCATTTTCGAC 804
QY 294 LeuLeuThrGluGluAspValAsp-----TyrAspAsnProGlyLeu 307
Db 805 CTTACGACGAACACACCGCCGTAACAGTTGATCGAAAAATACGTTTTCGCAACCG----- 858
QY 308 SerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTyr 327
Db 859 -----TTTATATTACAGAT-----GGTCAGTCCGAT 885
QY 328 IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGly 347
Db 886 CTGGCGCACTCAACGCAAGATATCATCTCAATAGCCAGTGCACGCCGCGCTTT----- 939
QY 348 PheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGly 367
Db 940 -----GATTACAGCTACGCCAG----- 957
```

```
QY 368 AspAlaValGluAsnPhePheAsnAsnGln----- 377
Db 958 -----GATAAATACAGGATAATCAGCGCGGTGTACCGCGTATGATGCACACGACA 1008
QY 378 ---ThrGlnAsnAlaArgIleGluLeu-----ArgHisGlnPro 389
Db 1009 GGAACACTGACACCGCGTGTGATGCAACTCAGGAGTCTACCCACGCGTATGCATGCTACT 1068
QY 390 IleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
Db 1069 CGTCGGGATCTGCAAGGGAAT-----GTTGATATTGCCGGAATCTATAATGACGATCTG 1122
QY 410 SerAlaThrSer-----GluAlaValLysGlnProMetLeuLeuAspAsnLys 425
Db 1123 GGTGGGTGTCATATGAATATTATGATCTTCTCGGTACAGATATGATTCGCTGTAAAAAA 1182
QY 426 ValGlnHisTyrSerPhePheGly----- 433
Db 1183 GCTAAAGATTTCATATATATACACCCCTGTTTATGTTAATACCAGCAAATGTACACCGTT 1242
QY 434 -----ValGluGlnAlaAsnTyr----- 439
Db 1243 TCGCGCTCGACACGCGATCAGACGATCAACAGAGAGAACTACTCAGCTTATGCACAGGAC 1302
QY 440 -----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLys 454
Db 1303 CGGCTCTATCTGACCGGATAACTGGATTGCCGTCGCCGGATCCGCTATCAGTATTACACG 1362
QY 455 AlaSerIleArgTyrAspLysAlaLeu---IleAspArgGluAsnTyrTyrLysGlnPro 473
Db 1363 CAATATCGGGTAAGCCGCTCTTTTAATGTCATTAATCTGACACGCCGCGATGAACAATGG 1422
QY 474 LeuProAspLeuGlyValAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyr 493
Db 1423 ACGCCCAAACTGGGGTAGCTAGTACAAACTGACGCCATCGGTATCTCTATTGGCCAATTAT 1482
QY 494 ---TyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeu 511
Db 1483 TCGCAAAACATTTATGCGCGAG-----TCGTCAATTGTCAGCTACATATGGCGATCTT 1533
QY 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 1534 CCACACAGAA-----TCATCTAATGCTTACGAAGTC 1563
QY 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
Db 1564 GGGCAGAAA-----TTCGAGCTATTTCGATGGTATCACC 1596
QY 552 GlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyr 571
Db 1597 GCAGAT-----ATTGCGCTGTTTGATATCCATATAAACGTAATGTGTGTAT 1641
QY 572 AlaGlnThrLeuAsnAsp-----GlyArgq---GlyProLysSer 583
Db 1642 ACCGAAAGTATTGTGTATGAACCATGCCAAAACGGCAGCGCGCTGCTTCAGAGGG 1701
QY 584 IleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db 1702 GTAGAAGTCGACCTTCGGGAGCATTAACTGAAACATTATAATCATATGCCAGCTACCGC 1761
QY 604 GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db 1762 TATACCGATGCAAGGTTCTGGAAGATCCT----- 1791
QY 624 AspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyr 643
Db 1792 GATTATGACGGGAACCATTTGCCGAATGTTCTCTCGTCATACCGGT----- 1836
QY 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 1837 ---TCGCTATTCTCGACCTATGATATTTCATTAACATGCCAGGCAATAACACTGACG 1890
QY 664 Leu-----GlyPheHis-----LeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
```

```
Db      1891  TTTGGCGGTGGGACATGTTGTAAGCGCGTTCGGCAACCAAT-----GGG 1938
Qy      680  LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
Db      1939  GCTGACTATTAT-----CTGCCT 1956
Qy      700  GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrp 719
Db      1957  GGCTATTTCGTTGCCGATGCCCTTCGCCGCATACAAAATGAAAATGCAGTAT---CCGGTC 2013
Qy      720  AsnTrpTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSer 737
Db      2014  ACTCTGCAATTAAACGTCAAAACCTGTTTGATATAAAGATTACACTCTTCC 2067

RESULT 17
US-10-085-959-152
; Sequence 152, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1e1 Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 152
; LENGTH: 7306
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-085-959-152

Alignment Scores:
Pred. No.:      4.15e-08      Length:      7306
Score:          177.50      Matches:      167
Percent Similarity: 31.47%      Conservative: 103
Best Local Similarity: 19.46%      Mismatches:  283
Query Match:      4.40%      Indels:      305
DB:              12          Gaps:        41

US-09-936-377-2 (1-758) x US-10-085-959-152 (1-7306)
Qy      6  LeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db      5128  CTCAAAGGTCGGCGCTCTTTTCACTGCTTTTTCGGCACCGATGATT-----CAT 5178
Qy      26  GluThrGluGlnSerValGlyLeuGlu-----ThrValThrValValGlyLys 41
Db      5179  GCACACAGACTGTGTAACGACCAAGATGGCGAAACAATCACTGTACAGCAGATGCAAA 5238
Qy      42  SerArgProAlaThrSerGlyLeu-----LeuHisThrSerThrAla----- 56
Db      5239  ACCGCACTGAGGCAACCGATGGTTATCAACCTCTGAGCACCTCCACGGGACATTAAAC 5298
Qy      57  -----SerAspLysIleIleSer 62
Db      5299  GATATGCCGATGTGGATATCCCGCAGGTGTCATACCGGTTAGCGATCAGGTTCTGGAA 5358
Qy      63  GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db      5359  -----AACCAAGATGGCAACACGCTGGATGAGCGGCTTTATAACGTCAGTAAC 5406
Qy      83  ---IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThr 101
Db      5407  GTGTTACAGCAACATACATTAGGGGGGACTCAGATGCTTTTGTACCCCGTGGTTGGC 5466
Qy      102  GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
```

```
Db      5467  GCRAACCGG----- 5475
Qy      122  ProAspHisAlaIleMetValAsp-----ThrAla 131
Db      5476  --GATGGCTCCATCATGACCAACACGGTCTGCGAACCGTACTTCTCGTAGTTTCAACGCC 5532
Qy      132  LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db      5533  GCAACAGAGCGTGTGGAAGTGTCTAAAGGCGCGCTCCACGCTGTATGGCATTTCTCGAT 5592
Qy      152  ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly 171
Db      5593  CCTGGCGGACTGATTAACTGCTGACCC---AAGCGCCCGGAAACAACTTCATCGTGTTCG 5649
Qy      172  ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGly 191
Db      5650  GTTTCAGCC-----ACCTCTCTCC 5667
Qy      192  GlyIleAsnIleGlyLeuGlyLys-----AsnPheValLeuHisThrGluGly 207
Db      5668  AGTTTTGGTGGCGGCACTGGGCAACTTGATATCACAGGTCCCATTTGAAGGCACTCAGCTG 5727
Qy      208  LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro 227
Db      5728  GCGTATCGCCTTACCGGGGAGTG----- 5751
Qy      228  AspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg 247
Db      5752  -----CAGGATCAA 5760
Qy      248  PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHis 267
Db      5761  GATTACTGGCGAAACTTCGGTAAAGCGCAGTACATTTATTGCCCCGTCA----- 5811
Qy      268  GluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArgTyr 287
Db      5812  -----CTCACCTGGTTTGGTGAT-----AATGCAACAGTA 5841
Qy      288  LeuGlnLeuTyrProHis----- 293
Db      5842  ACCATGCTCTATTCCCATCGGACTATAAACTCCATTTCGATCGTGGACGATTTTCGAC 5901
Qy      294  LeuLeuThrGluGluAspValAsp-----TyrAspAsnProGlyLeu 307
Db      5902  CTTACGACGAAACACAGCCCGTAAACGTTGATCGAAAAATACGTTTTCAGCAACCG----- 5955
Qy      308  SerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTrp 327
Db      5956  -----TTTAATATTACAGAT-----GGTCAGTCCGAT 5982
Qy      328  IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGly 347
Db      5983  CTGGCGCAACTCAACGCAAGATATCATCTCAATAGCCAGTGCACAGCGCGCTTT----- 6036
Qy      348  PheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGly 367
Db      6037  -----GATTACAGCTACAGCCAG----- 6054
Qy      368  AspAlaValGluAsnPhePheAsnAsnGln----- 377
Db      6055  -----GATAAATACAGCGATAATCAGCGCGTGTACCAGGTATGATGCAACGACA 6105
Qy      378  ---ThrGlnAsnAlaArgIleGluLeu-----ArgHisGlnPro 389
Db      6106  GGAACACTGACACGGCGTGTGTGATGCAACTCAGGGATCTACCACGCGTATGATGCTACT 6165
Qy      390  IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
Db      6166  CGTGGCGGATCTGAAGGGAAT-----GTTGATATTCCCGGATTTCTAATAGAGATTCTG 6219
Qy      410  SerAlaThrSer-----GluAlaValLysGlnProMetLeuLeuAspAsnLys 425
Db      6220  GGTGGGGGTGTCATATGATATTATGATCTTCTCGGTACAGATATGATTCGCTGTAAAAA 6279
```

```
QY 426 ValGlnHisTyrSerPheGly----- 433
Db 6280 GCTAAGATTTCATATATACACCCCTGTTATGGTAATACCAGCAATGTACACGGTT 6339
QY 434 -----ValGlnAlaAsnTrp----- 439
Db 6340 TCGCGCTCGGACGCGATCAGACGATCAAAAGAGAACTACTCAGCTTATGCACAGGAC 6399
QY 440 -----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLys 454
Db 6400 GCGCTCTATCGACCGATACTGATTCGCGTCCCGGATCCGCTATCATGATTATACAG 6459
QY 455 AlaSerIleArgTyrAspLysAlaLeu---IleAspArgGluAsnTyrTyrLysGlnPro 473
Db 6460 CAATATCGGTAAAGCGCCCTTTTAATGTCAATACTGACAGCCGCGATGAACAATGG 6519
QY 474 LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrp 493
Db 6520 ACGCCCAACTGGGGTTAGTCTACAAACTGACGCCATCGGTATCTTTATTTGCCAATAT 6579
QY 494 -----TyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeu 511
Db 6580 TCGCAACATTTATGCCGCG-TCGTCAATTCGACGCTACATTTGGCGATCTT 6630
QY 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 6631 CCACAGAA-----TCATCTAATGCTTACGAAGTC 6660
QY 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
Db 6661 GGGCAAAA-----TTCGAGCTATTCGATGGTATCAC 6693
QY 552 GlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyr 571
Db 6694 GCAGAT-----ATTGCGCTGTTGATATCCATAAAGCTAATGTGTGTAT 6738
QY 572 AlaGlnThrLeuAsnAsp-----GlyArg---GlyProLysSer 583
Db 6739 ACCGAAATATTTGTGATGAACCAATCGCCAAACCGGACGCGCGTTCGTTCAAGAGGG 6798
QY 584 IleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db 6799 GTAGAAGTCGACCTTGGGAGCAITTAACAGAAACATTATATCATTCGACGCTGCGC 6858
QY 604 GlyAlaGluGlyIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db 6859 TATACCGATGCAAGGTTCTGGAAGATCCT----- 6888
QY 624 AspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyr 643
Db 6889 GATTATGCGGGAACCAATTCGGAATGTTCTCGTCATACCGT----- 6933
QY 644 GlyAsnArgProPheIleAlaGlnAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 6934 -----TCGTATTCTCGACCTATGATTTATCAACATGCCAGCAATACACTGACG 6987
QY 664 Leu-----GlyPheHis-----LeuLysAlaSerLeuThrAspArgIleAspAlaAsn 679
Db 6988 TTTGGCGTGGGACATGGTAAAGCGCTGTTGCGCAACCAAT-----GGG 7035
QY 680 LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
Db 7036 GCTGACTATTAT-----CTGCCT 7053
QY 700 GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrp 719
Db 7054 GCCTATTTCGTTGCCGATCGCTTCGCCCATACAAATGAATTCGAGTAT---CCGCTC 7110
QY 720 AsnTrpTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSer 737
Db 7111 ACTCTGCAATTAAACGTCATAAACCCTGTTTGTATAAAACGTATTATACACCTCTTCC 7164
```

RESULT 18

```
US-10-238-075-261
; Sequence 261, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.B.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol.
; FILE REFERENCE: E.coli, and biological uses of these polynucleotides and of th.
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 261
; LENGTH: 7315
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-261

Alignment Scores:
Pred. No.: 4,16e-08 Length: 7315
Score: 177,50 Matches: 167
Percent Similarity: 31,47% Conservative: 103
Best Local Similarity: 19,46% Mismatches: 283
Query Match: 4,40% Indels: 305
DB: Gaps: 41
```

US-09-936-377-2 (1-758) x US-10-238-075-261 (1-7315)

```
QY 6 LeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGlnAlaHis 25
Db 5137 CTCAAAGGTGGCGCTCTTTTCACTGCTTTTTCGCGACCGATGATT-----CAT 5187
QY 26 GluThrGluGlnSerValGlyLeuGlu-----ThrValThrValValGlyLys 41
Db 5188 GCAACAGACTCTGTAACGACCAAGATGGCGAAACAATCACTGTACAGCAGATGCAAT 5247
QY 42 SerArgProArgAlaThrSerGlyLeu-----LeuHisThrSerThrAla----- 56
Db 5248 ACCGCAACTGAGGCAACCGATGGTTATCAACCTCTGAGCACCTCCACGGCGACATTAA 5307
QY 57 -----SerAspLysIleIleSer 62
Db 5308 GATATGCCGATGCTGGATATCCCGAGGTGCTCAATACGGTTAGCGATCAGGTTCTG 5367
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProgly 82
Db 5368 -----AACCAGATGCGACCAACGCTGGATGAGCGGCTTTATACGTCAGTAAC 5415
QY 83 ---IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThr 101
Db 5416 GTGTACAGACCAATACATAGGCGGACTCAGGATGCTTTTGTACGCGCGTGGTTTGGC 5475
QY 102 GlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSer 121
Db 5476 GCAAAACCGG----- 5484
QY 122 ProAspHisAlaIleMetValAsp-----ThrAla 131
Db 5485 ---GATGCTCCATCATGACCAACCGGTGCGAACCCTCTCTCTGCTAGTTTCAACGCC 5541
QY 132 LeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsn 151
Db 5542 GCAACAGACGCTGTGGAAGTGTAAAGGCCCGCCCTCCACGCTGATGGCATTCGAT 5601
QY 152 ValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGly 171
Db 5602 CCTGCGGACTGATTAACGCTGTCACC---AAGCGCGCGGAAAAACAATTCATCGTTGC 5658
QY 172 ValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlyLeuThrSerGly 191
```

Db 5659 GTTTCAGCC-----ACCTCCTCC 5676
Qy 192 GlyIleAsnIleGlyLeuGlyLys-----AsnPheValLeuHisThrGluGly 207
Db 5677 AGTTTGTGGCGGCACTGGCACTTGATATACAGGTCCCATTTGAAGGCACCTCAGCTG 5736
Qy 208 LeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuPro 227
Db 5737 GCGTATCGCCCTACCGGGAAGTG-----5760
Qy 228 AspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArg 247
Db 5761 -----CAGATGAA 5769
Qy 248 PheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSerHis 267
Db 5770 GATTACTGGCGAAACTTCGGTAAAGAGCGCAGTACATTTAATTCGCCCGTCA-----5820
Qy 268 GluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyr 287
Db 5821 -----CTCACCTGGTTTGGTGAT-----AATGCAACAGTA 5850
Qy 288 LeuGlnLeuTyrProHis-----293
Db 5951 ACCATGCTCTATTCCTACCGGACTATAAACTCCATTCGATCGTGGAAACGATTTTCGAC 5910
Qy 294 LeuLeuThrGluGluAspValasp-----TyrAspAsnProGlyLeu 307
Db 5911 CTTACGCAACAAACCGCGTAAACGTTGATCGAAAAATACGTTTTTGACGAACCG-----5964
Qy 308 SerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTyr 327
Db 5965 -----TTTAATATTACAGAT-----GGTCAGTCCGAT 5991
Qy 328 IleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGly 347
Db 5992 CTGGCGCAACTCAACGCAAGATATCATCTCAATAGCCAGTGGACAGCGCGCTTT-----6045
Qy 348 PheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGly 367
Db 6046 -----GATTACAGCTACAGCCAG-----6063
Qy 368 AspAlaValGluAsnPheAsnAsnGln-----377
Db 6064 -----GATAAATACAGCGATATCAGCGCGTGTACCGGTATGATGCAACAGCA 6114
Qy 378 ---ThrGlnAsnAlaArgIleGluLeu-----ArgHisGlnPro 389
Db 6115 GGAACACTGCACGCGGTGTGATGCAACTCAGGGATCTACCCAGCGTATGATGCTACT 6174
Qy 390 IleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
Db 6175 CGTGGGATGCAAGGGAAT-----GTTGATATGCGGATTCCTATAATGAGATTCTG 6228
Qy 410 SerAlaThrSer-----GluAlaValLysGlnProMetLeuLeuAspAsnLys 425
Db 6229 GGTGGGTGTCATATGATATATGATCTTCTGGGTACAGATATGATTGCTGTGTAATAA 6288
Qy 426 ValGlnHisTyrSerPhePheGly-----433
Db 6289 GCTAAAGATTCAATATATATACAACTGTTTATGGTAAATACAGCAAAATGTACAACGGTT 6348
Qy 434 -----ValGluGlnAlaAsnTyr-----439
Db 6349 TCGCGGTCCGACAGGATCAGACGATCAACAGGAGAACTACTCAGCTTATGCAACAGAC 6408
Qy 440 -----AspAsnPheThrLeuGluGlyValArgValGluLysGlnLys 454
Db 6409 GCGCTCTATCTGACCGATAATGATTGCTCCGTCGCGGATCCGCTATCAGTATTACAG 6468
Qy 455 AlaSerIleArgTyrAspPheAlaLeu---IleAspArgGluAsnTyrTyrLysGlnPro 473
Db 6469 CAATATCGGGTAAAGCGCGCTCTTTTAATGTCAATCTGACAGCCGCGATGAACAATGG 6528

Qy 474 LeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyr 493
Db 6529 ACGCCAAACTGGGTTAGTCTACAACTGACGCCATCGGTATCTTATTTGCCAATTAT 6588
Qy 494 -----TyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeu 511
Db 6589 TCGCAAAACATTATATGCCGAG-----TCGCAATTGGCAGCTACATATGGCGATCTT 6639
Qy 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 6640 CCACCAGNA-----TCATCTAATGCTTACGAAGTC 6669
Qy 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
Db 6670 GGGGCAAAA-----TTCGAGCTATTTCGATGATATCACC 6702
Qy 552 GlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyr 571
Db 6703 GCAGAT-----ATTGCGCTGTTTGATATCCATTAACGTAATGTGTGTAT 6747
Qy 572 AlaGlnThrLeuAsnAsp-----GlyArg-----GlyProLysSer 583
Db 6748 ACCGAAAGTATGTGTGATGAACCATCCCAAAACGGCAGCGCGTTCGTTCAGAGGG 6807
Qy 584 IleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyr 603
Db 6808 GTAGAAGTCGACCTTGGCGGAGCATTAACCTGAAACATTAAATATCATTTGCCAGCTACGCG 6867
Qy 604 GlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGly 623
Db 6868 TATACCGATCAAAAGGTTCTGGAAGATCCT-----6897
Qy 624 AspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyr 643
Db 6898 GATTATGCGAGGAACCATTCGCCGAATGTTCTCGTCATACCGGT-----6942
Qy 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 6943 -----TCGTATTCTCGACCTATGATATTCAATCATGCGCAGGCAATAACACTGACG 6996
Qy 664 Leu-----GlyPheHis-----LeuLysAlaSerLeuThrAspArgIleAspAlaSer 679
Db 6997 TTTGCGGTGGCGGACATGTTAAGCGCGTTCGTTCGGCAACCAAT-----GGG 7044
Qy 680 LeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrPro 699
Db 7045 GCTGACTATTAT-----CTGCTT 7062
Qy 700 GlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyr 719
Db 7063 GGCTATTTTCGTTGCGGATCGCTTCGCCGCATACAAAATGAAATTCAGTAT---CCGGTC 7119
Qy 720 AsnTyrTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSer 737
Db 7120 ACTCTGCAATTAAACGTCAAAACCTGTTTGTATAAACGATTACACCTCTTCC 7173

RESULT 19

US-10-085-959-24/c
; Sequence 24, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085.959
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1

SEQ ID NO 24
 LENGTH: 6732
 TYPE: DNA
 ORGANISM: Escherichia coli
 US-10-085-959-24

Alignment Scores:

Pred. No.: 4, 73e-08 Length: 6732
 Score: 176.50 Matches: 148
 Percent Similarity: 32.67% Conservative: 97
 Best Local Similarity: 19.73% Mismatches: 294
 Query Match: 4.37% Indels: 211
 DB: 32 Gaps: 32

US-09-936-377-2 (1-758) x US-10-085-959-24 (1-6732)

```

QY 27 ThrGluGlnSerValGlyLeu-----GluThrValThrValValGlyLysSerArg 43
D 1899 ACCAGAGTCCGTAGCGCTCAGAGAGTGACACGCTGACCGTATGGTCCAGTCCGGTA 1840
QY 44 ProArgAlaThrSerGlyLeuLeuHis-----ThrSerThrAlaSerAspLysIleile 61
D 1839 TCATCGACGACACACCGCTTCTCGATCAACCCACCATGAAGCCCTCGATAAA-----1786
QY 62 SerGlyAspThrLeuArgGlnLysAlaValAlaLeuGlyAspAlaLeuAspGlyValPro 81
D 1785 -----CAGAAATGCTCGCTCAGGCATTAAAGTGTCTGCCCC 1753
QY 82 GlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal---IleArgGlyGln 100
D 1752 GCGGTGGTGTCAAAAGTCAGGTAGCCGACAGGAAGAACAGGTCAAGTTCTGCTTT 1693
QY 101 ThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPhe 120
D 1692 GATATCGTCAGGTGCGGTCTATTTCGACGGTGTGCCCATTTATGCTTCCCTATCACGCG 1633
QY 121 SerProAspHisAlaIleMetValAspThrAlaLeuSerGlnValGluIleLeuArg 140
D 1632 AACCTGATCTGGCGGATCTGACCAACAATTTGGGG---GCAGTTGAAGTTTCCAAA 1576
QY 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspAlaAsp 160
D 1575 GGGTATTCGTCGCTCTCAGGGGCTAATCAGATGGGAGGCCATTAAATATCACACC 1516
QY 161 GlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgIleu 180
D 1515 CAGAAAG---CCGACTAAACCGCTGGAAGCCAGTCTGGGATATCGTCGGGATGGAAGCCGT 1459
QY 181 SerSerGlyAsnLeuGluLysLeu---ThrSerGlyGlyIleAsnIleGlyLeuGlyLys 199
D 1458 AGCCAGAACATGCCTACGATATGCTTCTATTGGCCGACGAGCATCTGGGG---1402
QY 200 AsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArg 219
D 1401 -----TATTTCGAAGTCAGCGGTAGCCAGCTAAAGCAGGATTTCTCGGCTGCGG---1351
QY 220 TyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArg 239
D 1350 -----CATGGTGTAAATAAT 1336
QY 240 AlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSerAspArgAspGln 259
D 1335 GATATTGAGGCAACACCGCAAGATGATTATTCATCGCTGATGATAACCGCGCAT 1276
QY 260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
D 1275 GTGAGCTGGTGTATACACCCAGCGAAACGATGAA-----1240
QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLysAsp 299
D 1239 -----TACATTCGACTTACATTAAGCAGGAT 1213
QY 300 ValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAla 319

```

```

D 1212 GGTGAAAAGATAAACCCTCATACAGT-----1186
QY 320 HisAlaHisAsnGlyLysProTyrPheAspLeuArgAsnLysArgTyrGluLeuArgAla 339
D 1185 ---GGAAATAGTGT-----CAGAAATCAGGTAC-----1159
QY 340 GluTyrLysGlnPro-----PheProGlyPheGluAlaLeu 351
D 1158 ---TGGCAGTGGCCAGAGTATGACAAAGAAAGTTTATTATCAGGAAACGACCACTA 1102
QY 352 ---ArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAla 369
D 1101 AACGATCGTTTACCTGAAAAAGTCGGCTGATCAC-----GACACC 1060
QY 370 ValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnPro 389
D 1059 TTTGAAATACGCTA-----ATGATGTACAACTCGCTGGGTGATTTGAAAAATAA---1009
QY 390 IleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeu 409
D 1008 -----AAAGGCAGCTACAGC---CATTTATCCGATTACAGCGACGTCGCGG 964
QY 410 SerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGln-----427
D 963 TTACAACCTGGCAGCGCATGTCGTGAAAAACGATCTGCTGTTTCCGCTTAACTGGAAA 904
QY 428 -----HisTyrSerPhePheGlyValGluGlnAla 437
D 903 GATGACGTACACCGGAAAAAGGTGCGCGGCACGCGCTTACGATCGCTATGAAGATCGT 844
QY 438 AsnTrp-----AspAsnPheThrLeuGluGly 446
D 843 ACCTGGTCGCTCGCCAGTGAATATCATATGGGCTGCTGCGGATAATGTCGATCGCTGGCT 784
QY 447 GlyVal-----ArgValGluLysGlnLysAla 455
D 783 GGAATCAGCTATGATGCGCGGATAGCTAGAGCGGAAAAACATGAGAGAGATGCGAGT 724
QY 456 SerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuPro 475
D 723 ATCAACCATTTATGACGAC-----706
QY 476 AspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTyrTyrPhe 495
D 705 -----AACAAATCAGTCAGCTTTTAACTGGCAGGTGATGGGAAAAATACCACTTT 658
QY 496 ThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGln 515
D 657 GCCAATGAAGCACCGCTGGCGCTTTCTGATATGACCGCACACGCTTTCCGACGCTGAAA 598
QY 516 GluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHis 535
D 597 GAAACGCTATACCGCTCCAAA---CCTGGTATACACAGATAGCGATGTTTAAACCGCAG 541
QY 536 LeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu---GlyTyrGluGlyAspArg 554
D 540 CTCGAACCGGAACCGCGCGCGGTGGATTAACTGGAATTTAAGCTGGTCTTCCGACGAC 481
QY 555 TrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThr 574
D 480 TGGGATTTGAGTCAGGCTTTTACTATAACCGGTGAGTATGATGCCATCTCTCGCAC---424
QY 575 LeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysValArg 594
D 423 -----AATATCGATGCGGATACC-----ATTCAAAAT 397
QY 595 TyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThr 614
D 396 CAGAACACGCGCACCGGTGATTACAGCGGTCTGGATGCGGTATTAAAGGGAAAAATCAGC 337
QY 615 ProArgTyrArgIleGlyValSer-----GlyAspTyrValArg-----627

```

Db 336 AATACTCGATGTAGGATTAGCTACGCCCTGATCCACGCTGATCCCAACGCAAAAGAC 277
QY 628 ---GlyArgLeuLysAsnLeuPro----- 634
Db 276 ATCGGCAAGATAACCGATCTGCCAAGCAGACAAATGACCGCATGGATGACTCTCAAACCG 217
QY 635 -----SerLeuProGlyArgGluAspAlaTyrGlyAsnArg 646
Db 216 TGGGAGCCGTTAAGCGTAACGCTGTGCGAAGAGCGCGTTCTCCAGCTACAGCAAC--- 160
QY 647 ProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeu----- 664
Db 159 -----AGTGACGGTTCACAAAAGCGCGGTTTTCGGGTGACCCACCATCGAGCC 109
QY 665 -----GlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAla 678
Db 108 GATTACACCTTAGTCTAGCTTCAGCGTTAATGTCGTCGTTCAATAAACCCTGTTGTATACC 49
QY 679 AsnLeuAspTyrTyrArgValPheAlaGln 688
Db 48 AATATGCTACAGTGAAGGGTTTATTGAA 19

RESULT 20
US-10-114-170-174
; Sequence 174, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Plunkett, Guy
; Perna, Nicole T.
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seav, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 174:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7304
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 174:

US-10-114-170-174

Alignment Scores:

Pred. No.: 1e-07 Length: 7304
Score: 174.00 Matches: 171
Percent Similarity: 31.41% Conservative: 101
Best Local Similarity: 19.75% Mismatches: 288
Query Match: 4.31% Indels: 306
DB: 14 Gaps: 42
US-09-936-377-2 (1-758) x US-10-114-170-174 (1-7304)
QY 1 MetAlaGlnThrThr-----LeuLysProIleValLeuSerIleLeuLeu 15
Db 5095 ATGGCTAAGTTACACCTTCATCTCAGGAATCAAGGTCGGGGCTCTTTTCACTGCTC 5154
QY 16 IleAsnThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThr 35
Db 5155 TTTGCAGCACCAATGATT---CATGCAACCGACACTGCAACGACCAAGATGCGGAACA 5211
QY 36 ValThrValVal-----GlyLysSerArgProArgAlaThrSerGlyLeu-----Leu 51
Db 5212 ATCACTGTTACAGCGGATGCAATACCGCAACTGAGGCGACCGATGGTTATCAACCTCTG 5271
QY 52 HisThrSerThrAla----- 56
Db 5272 AGCACTTCCACGGCGACATTAAACCGATATCCCGATCTCGATATCCCGAGGTGTCAT 5331
QY 57 -----SerAspLysIleIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGly 74
Db 5332 ACGGTTAGCGATCAGGTTCTGGAA-----AATCAGAATGCAACGACGCTGGAT 5379
QY 75 AspAlaLeuAspGlyValProGly---IleHisAlaSerGlnTyrGlyGlyAlaSer 93
Db 5380 GAAGCGCTTTATAACGTCAGTACGAGTACAGCAACCAATACATTAGCGGGAATCTAGGAC 5439
QY 94 AlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGlu 113
Db 5440 GCCTTTGTACGTCGTGTTTGTGCTAACCG----- 5472
QY 114 ThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAsp----- 129
Db 5473 -----GATGGCTCCATCATGACCAACGCGCTCGCAACT 5505
QY 130 -----ThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
Db 5506 GTACTTCCTCGCAGTTCACCGCCGACAGAAAGTGTGGAAGTGTAAAGGTTCGGCC 5565
QY 144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIle 163
Db 5566 TCCACGCTGTATGGCATTCTCGATCTCTGGTGGATTGATTAACTCGTGACC---AAGCGC 5622
QY 164 ProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly 183
Db 5623 CCGGAAAAAACATTCATGCTGTTCTGTCTCAGCC----- 5655
QY 184 AsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLys----- 199
Db 5656 -----ACGTCCTCCAGTTTGGCGAGGAGCTCGGCACTTGATATACACA 5700
QY 200 AsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArg 219
Db 5701 GGTCCCATTTGAAGGCACTCAGCTGGCATACCGCTGACGCGGGGAAGTG----- 5748
QY 220 TyrArgAsnLeuLysArgLeuProAspSerProArgPheAlaAsnGlyGlnHisArg 239
Db 5748 ----- 5748
QY 240 AlaValLeuGlyTrpArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGln 259
Db 5749 -----CAGGATGAAGATTACTGCGCAACTTCGGTAAAGAGCCGACGTACA 5793
QY 260 TyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTrpGln 279
Db 5794 TTTATTGCCCCGTCA-----CTCACCTGGTTT 5820

QY 280 LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHis----- 293
 Db 5821 GGTGAT-----AATGCAACAGTAACCATGCTCTATTCCCATCGGAGCATATAAACTCCG 5874
 QY 294 -----LeuLeuThrGluGluAspValAsp----- 301
 Db 5875 TTCGATCGTGGAAAGATTTTCGACCTTACACGAAACAGCCCGTAACAGTTGATCGAAAA 5934
 QY 302 -----TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 319
 Db 5935 ATACGTTTCGACGACG-----TTTAATATTACAGAT----- 5967
 QY 320 HisAlaHisAsnGlyLysProTyrPheAspLeuArgAsnLysArgTyrGluLeuArgAla 339
 Db 5968 -----GTCAGTCGATCTCGCGCAACTCAACGACAGATATCATCTCAATAGC 6015
 QY 340 GluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAsp 359
 Db 6016 CAGTGGACAGCGGCTTT-----GAT 6036
 QY 360 TyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGln----- 377
 Db 6037 TACAGCTACAGCCAG-----GATAAATACAGCGACAATCAGGCTCGC 6078
 QY 378 -----ThrGlnAsnAlaArgIleGluLeu----- 385
 Db 6079 GTTACCGCGTATGATCAACAGCAGCAACGCTGACCGCGGTGTGTGATGCAACTCAGGGA 6138
 QY 386 -----ArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyr 401
 Db 6139 TCTACCCAGCGTATGCAATCTACTCGTGGGATCTGCAAGGAAT-----GTTGATATT 6192
 QY 402 LeuGlyGlnLysSerAlaLeuSerAlaThrSer-----GluAlaValLys 417
 Db 6193 GCTGGGTCTATAATAGATTCTGGTGGGTGTCATATGAATATTATGATCTCTCGC 6252
 QY 418 GlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAla 437
 Db 6253 ACAGATATGATTCGCTGTAAACGCTAAAGATTCAATATCTACAAACCCCGTTTATGGC 6312
 QY 438 AsnTyrAspAsnPheThrLeu-----GluGlyGlyValArgValGluLysGln 453
 Db 6313 AATACCGCAATGCAACACGTTTCGGCGTCGACAGTATGATCAGATCAACACAGGAG 6372
 QY 454 LysAlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsn----- 468
 Db 6373 AGCTACTCAGCTTATGACAGAGATCGCTCTATCTGACCGATACTGGAATGCGCGTCCGC 6432
 QY 469 -----TyrTyr----- 470
 Db 6433 GGGATCGCTATCAGTATTATACCGCAGTATCGGGTAAAGCCGCTCTTTTAATGTCAAT 6492
 QY 471 -----LysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArg 485
 Db 6493 ACTGACAGCCGATGAACAAATGACGCCCAAACTGGGGTTAGTCTCAAACTGACGCCA 6552
 QY 486 SerPheAlaLeuSerGlyAsnTyr-----TyrPheThrProGlnHisLysLeuSerLeu 503
 Db 6553 TCGTATCTCTATTGCTAATATTATTCGCAACATTTATGCGCAA-----TCGTCA 6603
 QY 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
 Db 6604 ATTGCGAGTACATCGGAGATCTCCACCGAA----- 6636
 QY 524 ValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
 Db 6637 -----TCATCTAATGCTTACAAAGTCGGGCAAAA----- 6666
 QY 544 IleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArg 563
 Db 6667 TTCGAGCTGTTGATGGTATTAACCGCAGAT-----ATTGCGCTGTTGAT 6711
 QY 564 AsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAsp----- 577

Db 6712 ATCCATAAACGTAACGTTGTATACCGAAAGTATTGGTGATGAAACCATCGCAAAACG 6771
 QY 578 ---GlyArg---GlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr 595
 Db 6772 GCAGGCGCGTTCGTTCAAGAGGGTAGAAGTCACTTCGCGGAGCATTAACGAAAC 6831
 QY 596 AsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro 615
 Db 6832 ATTAATATCATTCGACGTACGCTATACCGCATTAAGGTCTGGAAGATCCT----- 6885
 QY 616 ArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSer 635
 Db 6886 -----GATTATGCGAGGAAACCATTCGCGAATGTTCTCCGT 6921
 QY 636 LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAspAspGlnAsn 655
 Db 6922 CATACCGGT-----TCGCTATTCTGACCTATGACATTTCATAAC 6960
 QY 656 AlaProArgValProAlaAlaArgLeu-----GlyPheHisLeu-----LysAlaSer 671
 Db 6961 ATCCAGGCAATAACACACTGACGTTTGGCGGTGGTGGACATTGCGTAAGCGCTCG 7020
 QY 672 LeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeu 691
 Db 7021 GCACCAAT-----GGGCGTGACTATTAT----- 7044
 QY 692 AlaArgTyrGluThrArgThrProGlyHisMetLeuAsnLeuGlyAlaAsnTyrArg 711
 Db 7045 -----CTGCCAGGCTATTTCGTTCCGATGCTTCGCGCATACAA 7086
 QY 712 ArgAsnThrArgTyrGlyGluTyrAsnTyrTyrValLysAlaAspAsnLeuAsnGln 731
 Db 7087 ATGAAATTCAGTAT---CCGTCACACTGCAATTAACCTCAAAACCTGTTGATAAA 7143
 QY 732 SerValTyrAlaHisSer 737
 Db 7144 ACGTATTACACCTCTTCC 7161
 RESULT 21
 US-10-181-319-13
 ; Sequence 13, Application US/10181319
 ; Publication No. US20030135032A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lewis, Thomas A.
 ; APPLICANT: Paszczynski, Andrzej
 ; APPLICANT: Crawford, Ronald L.
 ; APPLICANT: Cortese, Marc S.
 ; APPLICANT: Sebat, Jonathan L.
 ; TITLE OF INVENTION: Compositions and Methods for Bioremediation
 ; FILE REFERENCE: IDAHI19440
 ; CURRENT APPLICATION NUMBER: US/10/181,319
 ; CURRENT FILING DATE: 2002-07-15
 ; PRIOR APPLICATION NUMBER: PCT/US01/02386
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: 60/177,251
 ; PRIOR FILING DATE: 2000-01-20
 ; NUMBER OF SEQ ID NOS: 35
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 25801
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas stutzeri
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: 8300..8319, 8338..8708, 11735..11768, 18459, 18686, 18703, 18711, 18720, 19226, 2007
 ; OTHER INFORMATION: N = unknown nucleic acid residue
 US-10-181-319-13
 Alignment Scores:
 Pred. No.: 2,07e-06 Length: 25801
 Score: 169.50 Matches: 168
 Percent Similarity: 35.19% Conservative: 103

Best Local Similarity: 21.82%				Mismatches: 251			
Query Match: 4.20%				Indels: 253			
DB: 12				Gaps: 37			
US-09-936-377-2 (1-758) x US-10-181-319-13 (1-25801)							
QY	42	SerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIlelle	61	QY	350	-----AlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisaspGluLy	365
DB	12179	TCACAGCCGAGAAATCGAGCGCG-CTGAAAGGTGGCCCGCAGCGTGGCGGTATC	12237	DB	12921	TYCAGCATGAATTTGGCTCGGCTGGCATTCGGTAAACGGCTTGGAAACGATTCAAG	12980
QY	62	SerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGly--Val	80	QY	365	sAlaGly-----AspAlaValGluAsnPhenPheAsnAsnGlnThrGlnAs	380
DB	12238	GATGGCTGGACCGCAGCAGTCCAGCATCAGCTCAGCTCAACACTGGAGAGCGATT	12297	DB	12981	GACAGGATTCAGCAGGACACTGACTTCATCCAGCCGATGTTCTC-----TGCAGCTCGG	13034
QY	81	ProGlyIleHisAlaSerGlnTyrGly---GlyGlyAlaSerAlaProValIleArgGly	99	QY	380	nAlaArgIleGluLeuArgHisGlnPro-----IleGlyAr	392
DB	12298	CCTGGTCTGTCATTCCACAGCGCTTCGGGCAAGCAGGTATGATTAATCACCCTCATCGGGG	12357	DB	13035	CGCAGCATCACCTGGCGACACTCTCCAGAGGTTCGTGTGGAGGACAGCTCGGGAG	13094
QY	100	GlnThrGlyArgGlnLysValLeuAsnHisHisGlyGlnThrGlyAspMetAlaAsp	119	QY	392	gLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaTh	412
DB	12358	CTG-----ACGGCAACTTCAACAGC	12378	DB	13095	GCAGTTGGC---TGGCTGGTGTCTACGGGATCGCAGCAGCAACGATCTGCACAGTAC	13150
QY	120	PheSerProAspHisAlaIleMetValAsp-----	129	QY	412	rSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePh	432
DB	12379	TTCTCCAGTTCAAGTTGTTGCTGTCATGGCTTCCACGCTGCACAGCCAGGGATT	12438	QY	432	eGlyValGluGlnAlaAsnTrpAsp-----AsnPheThrLeuGluGlyGl	447
QY	130	-----ThrAlaLeuSerGln-GlnValGluIleLeuArgGlyProValThrLeuLe	146	DB	13204	CGCTGCACCTGTTACCCACTGGAACGTCCCTGTCGGCGGACTGGTCCATAGACGCGG	13263
DB	12439	CGAGAGTGGCATGCTGGATCTCGATCGATCGAGTTCATTCGCGGCCGCAATCTACGCT	12498	QY	447	yValArgValGluLysGlnLysAlaSerIleIleArgTyrAspLysAlaLeuLeuAspArgGl	467
QY	146	uTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIlePro---Gl	165	DB	13264	AGCGCGCTCGAGCGCAACGAGGTGCAGTACGTCCGCAAGGGGCTACG-----	13312
DB	12499	GTAATGGCGTAATGCCAGCGCGGTGTG-ATGGCATTCACAGCTTCGCGATGGA	12552	QY	467	uAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPh	487
QY	165	uLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg-----	179	DB	13313	-----AGCCATGAAAA-----	13324
DB	12553	CGCAGCCCGAGAGCCAGCGTGTCTGCCGAGCGGCGAGCGCGAAGCGGTGCATGCG	12612	QY	487	eAlaLeuSerGlyAsnTrpTyrPheThrPro-----GlnHisLysLeuSerLe	503
QY	180	-----LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyIleAsnIleGlyLe	197	DB	13325	-----GGWGGACACACAGTTTCCAGCGCTCGCGTGCACACACAG-----AT	13368
DB	12613	GTTTGGCTCAGCAGCGCTTGTGGAGAGRCGTTGTACGGCAGCGTA-----	12661	QY	503	uThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHi	523
QY	197	uGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaVa	217	DB	13369	AACGCCAATCACAATGGTATGTAGTGCAGTCTGGCGTGGCGCTGCGCGCTTCAA	13428
DB	12662	-----TCGGGCAACTGCTCG--	12676	QY	523	sValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAs	543
QY	217	lProArgTyrArgAsnLeuLysArgLeuProAspSerProArgPheAlaAsnGlyGl	237	DB	13429	TGTGCTGGCGCGAGCTG-----GGTTATCTGCTTACGACACGAGAGAAA	13476
DB	12677	-----AGCCAGAGCGGCTTCATCGACAACAC	12702	QY	543	nIleGluLeuAlaLeuGlyTyrGluGly-----AspArgTrpGlnTyrAsnLe	559
QY	237	nHisArgAlaValLeuGlyTrpArgLysArgPheTyrArgThrTyrSerAspArgAr	257	DB	13477	CTGTGCTGATGAACCGGTCTCAAGGCTGGCTTCTTGACAAGCGCATTCGCTATTGCT	13536
DB	12703	CCACAG-----	12709	QY	559	uAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyAr	579
QY	257	gAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIl	277	DB	13537	GGCGCGCTAC-----CTCATGGAC-----	13555
DB	12710	-----GGCACAAGCGGACGATCGTGAR-----	12733	QY	579	gGlyProLysSerIleGluAspSerSerGluMetLysLeu-----ValArgTy	595
QY	277	eTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGl	297	DB	13556	-----ATCATGACATGACAGGTTCATGACATGACATGACATGACATGACATG	13602
DB	12734	-----AACCTGGGCTGGCTGGGCGCGCGCGCGCAACGGA	12783	QY	595	rAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPr	615
QY	297	uGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAl	317	DB	13603	CATCAGCAGCGCTCCACGCGCAGCATCAAAGGTCTCGAGTGTGGACTA-TCTCC	13661
DB	12784	TGTGGTTCATGGCTAT-----	12801	QY	615	oArgTyrArgIleGlyValSerGly-----AspTy	625
QY	317	aHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGlu--	336	DB	13662	TCGTGGCGGCTGCACCTCAAGGCGGCGGCTGGCTGGAAACACACAGCGCGCTTCGATCA	13721
DB	12802	GCATCAGAGTAGTACAAATAGCGGCTCCCT-CTGGGGCTCGCCCGCGCGCGCAAGAAC	12860	QY	625	rValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAs	645
QY	337	-----LeuArgAlaGluTrpLysGlnProPheProGlyPheGlu-----	349	DB	13722	CTTTTCG-----GATGGCGAGCGGACTATGACGCAAA	13754
DB	12861	AAGTCCGTCGGAACCGCAACTGGAACCGTCTGTGAGGGCGCAGACCTTGTCTTCAATG	12920	QY	645	nArg-----ProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLe	664
				DB	13755	CCAGAACCGGTTTCGCGCC---GGATCTCACCGGCGCACCTTCGGCATCCGCTACGACGCGC	13811

664	Qy	U---GlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrT	583
13812	Db	CCGAAGCGCTGGTATGCACGAAGCAGCGTGGCAGCAGAGGTTCATCTGCATGCGG	13871
683	Qy	yrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisM	703
13872	Db	CCAAACGGGTATGAACCAAC-----GGCTACGGGCY	13901
703	Qy	etLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluThrAsnTprTyrV	723
13902	Db	TGCTGAACTTGGTAGTTGGTACCAACGC-----GGCAACTGGGMAATCGCGG	13949
723	Qy	alLysAlaAspAsnLeuLeuAsnGln	731
13950	Db	CCTACGCCCGCAACCGCGACCGATCAG	13975

RESULT 22

```

US-10-114-170-76/c
; Sequence 76, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
;             Burland, Valerie
;             Perna, Nicole T.
;             Flunkett, Guy
;             Welch, Rod
; TITLE OF INVENTION: NO. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Fincney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US

```

```
; ;  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word perfect

CURRENT APPLICATION DATA:

CORRECTION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/453,702

APPLICATION NUMBER: 03/4
FILING DATE: 03-DEC-1999

FILING DATE: 03-DEC-1955
 APPLICATION NUMBER: 60/110,955

; AFFILIATION NUMBER: 80/1
 ; FILING DATE: 04-DEC-1998

FILED DATE: 04-DEC-1993
ATTORNEY/AGENT INFORMATION:

NAME: Seav. Nicholas J.

NAME: Seay, Nicholas G.
REGISTRATION NUMBER: 27386

REGISTRATION NUMBER: 27388
REFERENCE/DOCKET NUMBER: 960296-95017; REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166

TELEFAX: (608) 251-9166
INFORMATION FOR SCO ID NO. 76.

; INFORMATION FOR SEQ ID NO: 76:
: SEQUENCE CHARACTERISTICS:
:

SEQUENCE CHARACTER;
LENGTH: 6737

```

;
;
; LENGTH: 6737
; TYPE: binary
;

```

TYPE: nucleic acid

STRANDEDNESS: double

TOPLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE

Alignment Scores:

Fragmented Occurs:	
Pred. No.:	9,78e-07
Score:	164.50
Percent Similarity:	33.02%
Best Local Similarity:	18.91%
Query Match:	4.08%
DB:	14
Gaps:	31
Indels:	213
Mismatches:	290
Conservative:	106
Matches:	142
Length:	6737

US-09-936-377-2 (1-758) x US-10-114-170-76 (1-6737)

	Qy	27	ThrGluGlnSerValGlyLeu-----GluThrValThrValValGlyLysSerArg	43
Db	1899	ACCACAGCCCGTAGCGGTCTCAGGAGAGCGCATCGTGCATGCCGTATGGTCACGTCGCGTA	1840	
	Qy	44	ProArgAlaThrSerGlyLeuLeuHis-----ThrSerThrAlaSerAspIlelle	61
Db	1839	TCATCGACGACGACCAACCGTTCTCGATCACCCACCATGAAGCCCTGGATAAA-----	1786	
Qy	62	SerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValPro	81	
Db	1785	-----CAGAATGTCGCTCAGGCATTAAAGTGTCGTCCTCCC	1753	
Qy	82	GlyIleHISalaserGlnTyrgLyglyAlaSerAlaProVal---IleArgGlyGln	100	
Db	1752	GGCGTGTGTCGTAANAAGTCAGCGCAGCGCAACGAAACAGGTAAAGTTTCGTGGCTTT	1693	
Qy	101	ThrGlyArggileLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaaspPhe	120	
Db	1692	GATAGTCGGCAGGTGCGGCTCTAATTCGACGGTGTGCCCATTTATGTTCCCTATGACGCG	1633	
Qy	121	SerProasPhisAlalleMetValAspThrAlaleuSerGlnGlnValGluilleuLeuArg	140	
Db	1632	AACCTCGATCTGGCGGGATCTTCACCAACAATCTGGG--CGACTCGAAGITTCACAA	1576	
Qy	141	GlyProValThrlenuTyrsrSerglyYasnValalaGlyLeuValaspValalaasp	160	
Db	1575	GGGTATTTCGTGCTCTTCAGGACCTAATCAGATGGCGCGACCAATTAATATCACCACC	1516	
Qy	161	GlylysileProglulysMetProgluasnglylval-----	172	
Db	1515	CAGAAG--CCAACAAAACCTCTGGAGCAAGTCTGGGATATGCCAGGAGTGGAGCGGT	1459	
Qy	173	-----SerGlyGiuLeugly---	177	
Db	1458	AGCCAGGACAATGCCTATGATATGCATGCTTCATTTGCCGCCAGACGCACTGGGGTAT	1399	
Qy	178	LeuArgLeusSrGlyYasnLeuGlululsLeuThrSergly-----GlyIleasn	194	
Db	1398	TTCAGCTAGCGGTAGCCAGCTAAAGCAGGATTTCTCGCCCTGCCGCGATGGTGAAT	1339	
Qy	195	IleglyLeungly-----LysasnPhe	201	
Db	1338	AATGATATTGCAGCAAACHCGCAAGATGATTAATTATCGGCTGATGATAAACGGCGC	1279	
Qy	202	ValLeuHisThrGliuGlyLeuTyraArgLysSerglyAspTyralaValProArgTyraArg	221	
Db	1278	ATTCTGAAGCTCGCATTTACACCACGTGAAACCGATGAATACACACTGACTTACATTAA	1219	
Qy	222	AsnLeuylsArgLeuProaspSerProargPhealaenglyGlnHisArgAlaval	241	
Db	1218	CAGGATGGTGA AAAAGATAACCCGCCATCAGCGGAAATAGTGTGCTAAAAATCAGCTAC	1159	
Qy	242	LeuglyTrp-----ArgLysArphetyrArgthrTyrsrAspArgArg	257	
Db	1158	TGGCAGTGGCCAGATATGACAAGAAGTTTTATTATCAGGGAACGACCAACTAAAC	1099	
Qy	258	AspGlnTyrglyLeuproalaHisserHiselutyrAspAspCysHisAlaaspIlefile	277	
Db	1098	GATCGTTTTACCCTGAAAAGTCGGCTGTATCGC--GACACCTTTGAAAAATACGCTGAT	1042	
Qy	278	TrpGlnylsSerleulleasn--LysArgTyrlenuLeuTyrrProHisLeuleuthr	296	
Db	1041	ATGPACAACCTCGCTGGCTGATTTGAAAAATAAAAAGCGAGCTACAGCCATTAT-	988	
Qy	297	GluguaspValAsptyr---AspasnProgllyLeuSerCysGlyPheHisAspAspAsp	315	
Db	987	-----TCCGATTACAGCGAGGTGCCGATTACAACCTGGCA-----	952	
Qy	316	AspalahisalalahiasnglylaseProtripleaspLeuArg---AsnLysArg	334	

D	b	951	----	GCGGATGTCGGTGAACCATCTG	922
Q	y	335	TyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHis	354	
D	b	927	CTGACGTTTGCCGTAACTCGAAA	904	
Q	y	355	LeuAsnArgAsnAspTyrHisHisaspGluLysAlaGlyAspAlaValcIuasnPhePhe	374	
D	b	903	-----GATGACGTACACCGGAAAAAGTGC CGCAGCCGCTTACGATCGCTAT	853	
Q	y	375	AsnAsnGlnThrGlnAsnAlaAargileGluLeuArgHisGlnProileGlyArgLeuLys	394	
D	b	852	GAAGATCGTACTGCTCGCCAGTGAATCAATGCGCTGCTCGCGATAATGTCGAT	793	
Q	y	395	GlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGlu	414	
D	b	792	GTTGTGGCTGGAATCAGCTATGACTACGGCGCATAGC-----GTAGAA	751	
Q	y	415	AlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyVal	434	
D	b	750	GCGRAAAAACACGAGAAA-----GATGCGCACATCACCAATTATGAC-----	709	
Q	y	435	GluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLys	454	
D	b	708	-----GACAAC	703	
Q	y	455	AlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeu	474	
D	b	703	-----	703	
Q	y	475	ProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyr	494	
D	b	702	-----AATCAGTCAGCTTTTACTGGCAGCTGATGCGGAAATACCAC	661	
Q	y	495	PheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThr	514	
D	b	660	TTTGTCAATGAAGACACGCTGCGCTTTCGTACTATGACCGCAAACGCTTTCGAGCGTG	601	
Q	y	515	GlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys	534	
D	b	600	AAAGAAGCTATACCATCCACATCCAAA--CCTGGGTATAACAGATACGATTTGTTAACCCG	544	
Q	y	535	HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu---GlyTyrGluGlyAsp	553	
D	b	543	CAGCTCAACCGCGAAACCGCTCGCGGGTGGATTTAACTGGAATGGTCCCTTCACGCAC	484	
Q	y	554	ArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGln	573	
D	b	483	GACTGGGGATTGAGTTCAGCGCTTTACTATAACCGGGTGATGCCATCCTCTCGCAC	424	
Q	y	574	ThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuVal	593	
D	b	423	-----AATATCGATCGCGATACC-----ATTCAA	400	
Q	y	594	ArgTyrAsnGlnSerGlyAlaAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysPro	613	
D	b	399	AATCAGACAGCGGCACCGGTGATTCACCGCTCTGGATCCCGGTATTAAAGGGAAATC	340	
Q	y	614	ThrProArgTyrArgIleGlyValSer-----GlyAspTyrValArg---	627	
D	b	339	AGCAATATACTGGATTAGATTGATGCTACGCCCTGTATCCAGCTGACGCCAAACGTAA	280	
Q	y	628	-----GlyArgLeuLysAsnLeuPro-----	634	
D	b	279	GACATCGGCAAGATACCGATTGCCCCACGACACATGACCGCATGGATGACTCTCAA	220	
Q	y	635	-----SerLeuProGlyArgGluAspAlaTyrGlyAsn	645	
D	b	219	CCGTGGAGCCGTTAAGCGTAAACGCTGTCGGAAGAGCGCGTTCCTCCAGCTACAGTAC	160	
Q	y	646	ArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlalaArgLeu---	664	
D	b	159	-----AGTGACGGTTTCAAAAAGCCGCGGTTTTGCGGTGACCCATCGA	112	

Alignment Scores:	
Pred. No.:	3.75e-05
Length:	76804

Db 55388 ---ACCATCAGTGTGAAGACGACAGCGCCCGCTATTTCACGGCTGGAAGTGCAGTGGAC 55332
QY 610 TyrPheLysProThrProArgTyrArgIleGly----- 620
Db 55331 TACCTGATCCGGATACCTGACTGAGTACCGGTGCTGNACTTCAATGTGCTGAAACCGAG 55272
QY 621 -----ValSerGlyAspTyrValArgGlyArgLeuLysAsnLeuProSerLeuProGly 638
Db 55271 TCGAAGTGAACGGTCAATGGCAAAATATGACGTGAAGAA-----TCAAGTCCATCG 55218
QY 639 ArgGluAspAlaTyr----- 643
Db 55217 AAAGCGACAGCTTACATTAACTAGGCGCCCGCAACCGTGGAGTCTGGTGTACAGAGACC 55158
QY 644 -----GlyAsn-----ArgProPheIleAlaGlnAsp 652
Db 55157 ACTTCTTCGACGTAAAGCATGACGAGGGTAACGATATTAAATGGTTACACTACCGTGCAT 55098
QY 653 AspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeu 672
Db 55097 TTTATCAGTAGTTGGCAGCTTCGGGTGGGAACACTCAGCTTCAGCGTTGAG---AACCTC 55041
QY 673 ThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAla 692
Db 55040 TTGACCGT-----GACTATACCACTGTCTGGGACAGCGTGCACCTCTG 54996
QY 693 ArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeu----- 706
Db 54995 TACTAC-----AGCCCGGGTTACGGCCCTGCTTCACTGTACGACTACAAAGGCGG 54945
QY 707 -----GlyAlaAsnTyr 710
Db 54944 GGCGAAGCTTTGGTGTGAAGTAC 54921

RESULT 24

US-10-043-344-3

; Sequence 3, Application US/10043344

; Publication No. US20030088086A1

; GENERAL INFORMATION:

; APPLICANT: Loesmore, Sheena M.

; APPLICANT: Harkness, Robin E.

; APPLICANT: Schryvers, Anthony B.

; APPLICANT: Chong, Pele

; APPLICANT: Gray-Owen, Scott

; APPLICANT: Murdin, Andrew D.

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES

; FILE REFERENCE: 1038-1221 MIS

; CURRENT APPLICATION NUMBER: US/10/043,344

; PRIOR FILING DATE: 2002-07-01

; PRIOR FILING DATE: 1996-05-17

; NUMBER OF SEQ ID NOS: 160

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 5009

; TYPE: DNA

; ORGANISM: Haemophilus influenzae

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (121)..(2100)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (2117)..(4852)

; US-10-043-344-3

Alignment Scores:

Pred. No.:	1,34e-06	Length:	5009
Score:	161.50	Matches:	199
Percent Similarity:	32.32%	Conservative:	130
Best Local Similarity:	19.55%	Mismatches:	318
Query Match:	4.00%	Indels:	371
DB:	14	Gaps:	54

US-09-936-377-2 (1-758) x US-10-043-344-3 (1-5009)
QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db 2120 ACTAAAAAACCCCTATTTCGCTAAGTATTATTCTTGTCTTTAAATTCATGCTATGTA 2179
QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
Db 2180 AAAGCAGAAACTCAAAAGTATAAAAGATACAAAAGAGACTATATCATCTGAAGTGGACACT 2239
QY 33 -----LeuGluThrValThrValValGly-----LysSerArgPro 44
Db 2240 CAAAGTACAGAAAGATTAGAAACTATTCTCAGTCACTGCAGAAAATAAAGAGAT 2299
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2300 CGTAAAGATAATGAAGTA-----ACTGGACTTGGCAAAATATTCAAAACCTAGT 2347
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2348 GAAAGTATCAGCGGAGAACCAAGTATTAAATATTTCGTGATCTAACACGCTATGATCCAGG 2407
QY 83 IleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProVal-----IleArgGlyGlnThr 101
Db 2408 ATTTCAGTTGTAGAACCAAGTTCGGGTGCAAGTCTCGGATATTCTATTTCGGTATGGAC 2467
QY 102 GlyArgArgIleLysValLeu----- 108
Db 2468 AGAAATAGAGTTCGTTTATTAGTAGTGGTTTACCTCAAAGCAATCTTATGTAGTCAA 2527
QY 109 -----AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro 122
Db 2528 AGCCCTTTAGTTCGCTCGTTCAGGATATTCTGGCATATTCTGTCATTAATGAATGAATAT 2587
QY 123 AspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyPro 142
Db 2588 GAAATGTA-----AAGGCGGTGCAAAATGAAGCAAGGGGGG 2623
QY 143 ValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal----- 156
Db 2624 AGTCTCTCTGAGTATGCTAATGAGCAGTACTAGTCTGTCTGTAACTTTCAAAGCAATCA 2683
QY 157 -----AspValAlaAspGly-----LysIleProGluLys 166
Db 2684 GCAGCCGATATCTTAGAAGGAGACAAATCATCGGGAATTCAAACTTAAATATGCTTATTC 2743
QY 167 MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186
Db 2744 AGCAAAATAAAGGCTTTTACCCTTTTACCTGTAGCAGGAAAAACAAGGTGGATTGAA 2803
QY 187 LysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheVal-----Leu 203
Db 2804 -----GGCTAGCCATTTCACCTCAACGAAATTCAAATGAAACCCCAAGTC 2848
QY 204 HisThrGluGlyLeu-----TyrArgLys 211
Db 2849 CATAAAGATGCATTTAAAGGCGGTACAAAGTTATGATCGATTAAATCGCCACACAGATAAA 2908
QY 212 SerGlyAspTyr----- 215
Db 2909 TCCTCAGGATACTTTGTGATACAAAGGTGAGTCCAAATGGTGTATGACAAAGTGTGCAGCC 2968
QY 216 -----AlaValProArgTyrArgAsn 222
Db 2969 AAGCCACCTCGACTTTTCCACCAAGGAAACCGTAAGCGTTTCAGATTATACGGGG 3028
QY 223 LeuLysArgLeuProAspSerProArgArgPheAlaAsn----- 235
Db 3029 GCTAACCGTATCAACCTAATCCATGAAATGAAAGCCAGTCTTGGTTTTTAAGAGGA 3088
QY 236 GlyGlnHis-----ArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThr 252
Db 236 GlyGlnHis-----ArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThr 252

Db 3089 GGGTATCATTTTCTGAACAACATTATATTTGGTATTTTTTGAATTCACACACAAAA 3148
Qy 253 Tyr-----SerAspArgAsp--- 258
Db 3149 TTTGATATCCGTGATATGACATTTCCCGCTTATTTAGCCCAACAGAAAGCGGATGAT 3208
Qy 259 -----GlnTyrGlyLeuProAlaHisSerHisGlu--- 268
Db 3209 AGTAGTCGTTCTTTTATCCAATCAAGATCATGGTGCATATCAACATATTTAGGATGGC 3268
Qy 269 -----TyrAspAspCysHis----- 273
Db 3269 AGAGCGGTAAATATGCAAGTGGGCTTTATTTTCATGTAACACCATAGAGAAACAACGGTGTA 3328
Qy 274 -----AlaAspIleIle----- 277
Db 3329 GGTATTGAATATATTTACGAATAAAGAACAAAGCGGCATCATTTGACAAAGCAGTGTTA 3388
Qy 278 -----TrpGlnLysSerLeuIleAsnLysArgTyrLeuGln-----Leu 290
Db 3389 AGTGCTAATCAACAAACATCATCTTGACAGTTATATGCGACATATCGCATTTGCAGTCTT 3448
Qy 291 TyrProHisLeuLeuThrGluGluaspValaspTyrAspAsnProGlyLeuSerCysGly 310
Db 3449 TATCCT-----AATCCAAGTAAGAATTGCCGC 3475
Qy 311 -----PheHisAspAspAspAlaHisAlaHisAla 321
Db 3476 CCAACACTTGATAAACCTTATTCATCTATCTGATAGAAATGTTTATTAAGAAAAA 3535
Qy 322 HisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrp 341
Db 3536 CATAAT-----ATGTTGCAATTG---AATTTAGAGAAAAAATTCACAAAAATTGG 3583
Qy 342 -----LysGlnProPhePro---GlyPheGluAlaLeuArgValHisLeuAsnArg 357
Db 3584 CTTACTCATCAATGTTCTTCAATCTTGTTTGTGATGACTTACTTCAGCGCTTCAGCAT 3643
Qy 358 AsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPheAsnAsnGln 377
Db 3644 AAGATTAT----- 3652
Qy 378 ThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrp 397
Db 3653 ---TTAACTCGAGTGTATCGCTACGCGCAGATAGTATTTCCAAGGAAACCTGGTGAAACT 3709
Qy 398 GlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLys 417
Db 3710 GGTAAACCAAGAAATGGTTGCAATCA----- 3736
Qy 418 GlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAla 437
Db 3737 CAACCTTACTTTATACCCA---AAACAGCAGCCATAT---TTTGCAGGACAAGATCATTTGT 3790
Qy 438 AsnTrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIle 457
Db 3791 AATTATCAA-----GGTAGCTCTCTAATTACAGACACTGTAAGTG 3832
Qy 458 ArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyr----- 470
Db 3833 CGG-----TTAATTAAAGGMAAAATTTATTTTCGCGACGCGCAATAATATG 3880
Qy 471 ---LysGlnProLeuProAspLeuGly----- 478
Db 3881 GCATTAGGGAATAACGTTGATTTAGGTTTTCGATATCGGTATCGATATCTCGTACAAAA 3940
Qy 479 AlaHisArgGlnThrAla-----ArgSerPheAlaLeuSerGlyAsnTrp 493
Db 3941 GTATAGATCAACTATAGTGTGGTAAATTTAAAAATTTCTTGGAAATCTGGTATT 4000
Qy 494 TyrPheThrProGlnHisLysLeuSerLeuThr-----AlaSerHisGlnGluArgLeu 511
Db 4001 GTCATAAAACCAACGAAGTGTGATCTTTCTTATTCGCGCTTCTACTGGATTAGAAAT 4060

Qy 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
Db 4061 CCTAGTTTTTCTGAAATGATAT-----GGTTGGCGGTATGGTGGCAAAATGACAGGTT 4114
Qy 532 GlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGlu 551
Db 4115 TATGTAGTAAATTTAAAGCCCTGAAACATCTCGTAACCAAGAGTTTGGTCTCGCTCTAAAA 4174
Qy 552 GlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIle--- 570
Db 4175 GGGGATTTTGGTATATATGAGATCAGTCATTTTATAGTAATGCTTATCGAAATCTTATCGCC 4234
Qy 571 TyrAlaGlnThrLeuAsn-----AspGlyArgGlyProLysSerIleGluAspAsp 587
Db 4235 TTTCTCGAAGAACTTAGTAAAAATGGAACCTGGAAGGGCAATTATGGATATCATAAATGCA 4294
Qy 588 SerGluMetLysLeuValArgTyrAsn---GlnSerGlyAlaAspPheTyrGly----- 604
Db 4295 CAAAATGCAAAATTAGTTGGCGTAAATATACTCCACAATTAGATTTTAAATGGTTATGG 4354
Qy 605 -----AlaGluGlyGluIleTyrPheLys--- 612
Db 4355 AAACGTATTCCTACGTTGGTATGCAACATTTGCTTATACCAAGTAAAGCTTAAAGAT 4414
Qy 613 -----ProThr 614
Db 4415 CAAAAATCAATGCTGTTAGCCTCCGTAAGCAGTTATTTATTTGATGCCATTCAGGCC 4474
Qy 615 ProArgTyrArgIleGlyValSerGlyAsp----- 624
Db 4475 AGCGGTATATCATTTAGTTAGGTATGATCATCCAGTAATATCTTTGGGGAATTAATACA 4534
Qy 625 ---TyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyr 643
Db 4535 ATGTTTACTCAATCAAAAGCAAAATCTCAAAATGAATTTGCTAGGAAAACGT---GCATTA 4591
Qy 644 GlyAsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArg 663
Db 4592 GGTAACAAT-----TCAAGGATGTAATAATCAACAAGAAAACCTTACTCGGCA--- 4639
Qy 664 LeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyr 683
Db 4640 -----TGCATATC-----TTAGATGTATCG 4660
Qy 684 ArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMet 703
Db 4661 GGTATTATACATGCGCAATAAA-----AATATTATG 4690
Qy 704 LeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyLeuTrpAsnTrpTyrVal 723
Db 4691 CTTTCGATTAGGATATATAATTTTATTCAACTATCGCTATGTTTACTTGGAAACGGTGGCT 4750
Qy 724 Lys---AlaAspAsnLeuLeuAsn-----GlnSerValTyrAlaHisSerSerPheLeu 740
Db 4751 CAAACAGCACAGGTCGGTCAATCAACATCAAAATGTTGTAGTACTACTCGTACGCA 4810
Qy 741 SerAspThrProGlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
Db 4811 GCATCA-----GGACGAACATATACCTTAACATTAGAAAATGAAATTC 4852

RESULT 25

US-10-043-344-2

; Sequence 2, Application US/10043344

; Publication No. US2003008086A1

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Harkness, Robin E.

; APPLICANT: Schryvers, Anthony B.

; APPLICANT: Chong, Pele

; APPLICANT: Gray-Owen, Scott

; APPLICANT: Mordin, Andrew D.

; APPLICANT: Klein, Michel H.

Qy	5	ThrLeuIysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu	20
Db	2168	ACTAAACAAACCCATTATTTCCGCTAAGTATTATTTCTGTCTTTTAAATTCATGCTATGTA	2227
Qy	21	LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly-----	32
Db	2228	AAMCGAGAACTCAAAGTATAAAAGTATACAAAGAGCTATATCATCTGAAGTGGACACT	2287
Qy	33	-----LeuGluThrValThrValValGly-----LysSerArgPro	44
Db	2288	CAAAAGTACAGAGATTCAGAAATTAGAAACTATCTCAGTCACTGCAGAAAAAATAAGAGAT	2347
Qy	45	AcqAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser	62
Db	2348	CGFAAAGATAAATGAAGTA-----ACTGGACTTGGCAAAATATCAAAACTAGT	2395
Qy	63	GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly	82
Db	2396	GAAGATATCAGCCGAGAACAGTATAAATATTCGTGATCTAACACGCTATGATCCAGGG	2455
Qy	83	IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr	101
Db	2456	ATTTCAGTTGTAGAACAAAGTCGCGGTGCAAGTCTCGATATTTCTATTCGTGTATGGAC	2515
Qy	102	GlyArgArgIleLysValLeu-----Lys-----	108
Db	2516	AGAAATAGAGTTGCTTTATTAGTAGTGGTTTACCTCAAACGCNACTTATGTAGTGCAA	2575
Qy	109	-----AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro	122
Db	2576	AGCCCTTTAGTTGCTTCGTTCAGGATATTTCCGCAATTAATGAATTTGAATAT	2635
Qy	123	AspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyPro	142
Db	2636	GAATAATGTA-----AAGCCCTGCGAAATAAACAAGGGGGGG	2671
Qy	143	ValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal-----	156
Db	2672	AGTCTCTGAGTATGGTAATGGAGCACTAGCTGGTCTGTACATCTTCAAGCAANTCA	2731
Qy	157	-----AspValAlaAspGly-----LysIleProGluLys	166
Db	2732	GCAGCCGATATCTTAGAAGAGAGACAAATCATGGGAATTCAAACTAAAAATGCTTATCA	2791

US-09-936-377-2 (1-758) x US-10-114-170-119 (1-3275)

```
QY 209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys-----ArgLeu 226
D 103 TATCGCATCAACGGTAGTATTATTCATCATAGGCAATCGTGATACCGCGGATGGAGCTCTG 162
QY 227 ProAspSerProArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLys 246
D 163 CCGATATACC-----AACTATCGTAAACAATAGTACAGGTGTATGGTGTATTAACCTCC 216
QY 247 ArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSer 266
D 217 GGAACCATCGT-----TTTGGCCTCTCGTGTATCGCTACAGACTCCGACACAACCT 270
QY 267 HisGluTyrAspAspCysHisAlaAspLeuIleTrpGlnLysSerLeuIleAsnLysArg 286
D 271 TAC--TATGAGGATCCAGACGGAATTTAT-----GAGGCA 303
QY 287 TyrLeuGlnLeuTyrProHisLeuLeuThrGluLys-----AspValAsp 301
D 304 TTTAGTGTCAAAATACCTAACTTGAACGAGAGAAGTTGGGTATTTCTATGACACAGAC 363
QY 302 TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAla 321
D 364 GTGACGGTGACTATCTAAAA-----AAAATTCTATTCGACGGC 402
QY 322 HisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrp 341
D 403 TATGACGACACATCCAGCGCAATTTGCCAAC-----GAAGTAAAAACGACACAG 453
QY 342 LysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis 361
D 454 CTGTGTTCCAGTCCGATGATTCAGGTCTGACGGTTCAT---AACAGACTGAC---ACC 507
QY 362 HisAspGluLysAlaGlyAspAlaVal-----GluAsnPhePheAsnAsnGlnThrGln 379
D 508 CATGATAAGCAATACACTCAGCGGTGCACATTCGACAGTCACTTTTCGCTGCTGCTAAT 567
QY 380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
D 568 AATGAACCTGTTTACC-----GGTGCA 588
QY 400 GlnTyr-----LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla 415
D 589 CAGTACAACAAGATAGGTGACCAAGGTCCGGTGGCATGACCTCAAGCAAAATCTCTG 648
QY 416 ValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlu 435
D 649 ACCGGC-----TTCATTATAAGGAACACGAACTCGCTCTATTATGATCAGTCAAG 699
QY 436 GlnAla-----AsnTrp-----AspAsnPheThrLeu 444
D 700 CAAGTACAGTCTCACTATTCGCACAAATGACTGGCAATTCGCCGATCACTGGACATGG 759
QY 445 GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle 464
D 760 ACAATGGAGTTCCG-----CAATACTGGCTTCTTCAAGTTGACGGTGT-----807
QY 465 AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla 484
D 808 GACGGAGTATCATATACCGCAGCAATTAAGCGATACCTCTCTCCAGAGAGTCTGCG 867
QY 485 -----ArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer 502
D 868 AGTGATCAGAATGTGTAACATCTACAGCTCGCTATTTCCAGTTTCAGTTTCGATTAACCTGGAG 927
QY 503 LeuThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuTyrAlaHis 520
D 928 TTAGCGGTGCGTTCGCGCAAGCTAGTATTTCCACACTCTCCCGAGCTTTTATGCGAG 987
QY 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
D 988 ACATCTCGCGGGCGGCGAGTGTCAATAC-----GGAATCTCTGATCTTAAGGCTGAACAC 1041
```

```
QY 541 SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAla 560
D 1042 TCCAATAACTTTGAATTAGGTGCAGATATAATGTTAATCAGTGGCTGATGACACGCCA 1101
QY 561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
D 1102 GTTTACTACTCAGACGTAAAGATTATATT---GCAAGTCTGATCTGTGATGGCAGTATA 1158
QY 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr-----AsnGlnSer 598
D 1159 GTTTCGAATGGTAACACCAACTCTCCCGTAGCTAGTACTATTATTATGACAATAATTGAT 1218
QY 599 GlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArg 618
D 1219 CGGGCAAAACATCGGGACTGGAA-----ATAAGCGCGGAATAATAAT 1260
QY 619 IleGlyValSerGlyAspTyrValArgGlyArgLeu-----630
D 1261 GGCTGGTTTTCTCGCATATATCAGTGGCAATTTAATTCGTGCGCAATATGAACTTCA 1320
QY 631 -----LysAsnLeuProSerLeuProGlyArg-----639
D 1321 ACATTAAAAACAACATAACAGTGAACCCAGCGATAAACGGACGTATAGGGCTGAAACAT 1380
QY 640 -----GluAspAlaTyrGlyAsnArgPropheIle 649
D 1381 ACTCTGTGATGGTCCAGCCCAACATACTCTGATGTTTATTTCGTGCTGCCTCTAGT 1440
QY 650 AlaGlnAspAspGlnAsnAla-----ProArgValProAla---AlaArgLeuGlyPhe 666
D 1441 GCAAAAGATGACAGTAAACGGTACCGGAAACAATGTTCCGGCTGGGCCACTCTCAACTTT 1500
QY 667 HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal---685
D 1501 GCAGTAAATACAGAAATTC-----GGTAACGAGGATCAGTACCGGATTAAC 1545
QY 686 PheAlaGlnAsnLysLeuAla-----ArgTyrGluThr-----ArgThrProGly 700
D 1546 CTGGCAGCTCAATAACCTGACAGACAAACGCTACCGTACAGCACATGAAACTATTCTCTGCA 1605
QY 701 HisHisMetLeuAsnLeu-----GlyAlaAsnTyrArgArgAsnThrArgTyr 716
D 1606 GCAGGTTTTTATTCAGCTATAGTGTGTTGATGGAATTTCTGATGACAAAAACAATCGTTA 1665
QY 716 rGly 717
D 1666 TGA 1669
```

RESULT 27

US-10-238-075-719

; Sequence 719, Application US/10238075

; Publication No. US20030148324A1

; GENERAL INFORMATION:

; APPLICANT: I.N.S.E.R.M.

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D- A- and which are isola

; FILE REFERENCE: BLANDINE

; CURRENT APPLICATION NUMBER: US/10/238,075

; PRIOR FILING DATE: 2002-09-10

; PRIOR APPLICATION NUMBER: 0003145

; NUMBER OF SEQ ID NOS: 1576

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 719

; LENGTH: 5585

; TYPE: DNA

; ORGANISM: Escherichia coli

US-10-238-075-719

Alignment Scores:

Pred. No.: 2,03e-06 Length: 5585

Score: 160.50 Matches: 129

Percent Similarity: 36.25% Conservative: 82
 Best Local Similarity: 22.16% Mismatches: 239
 Query Match: 3.98% Indels: 132
 DB: 12 Gaps: 33

US-09-936-377-2 (1-758) x US-10-238-075-719 (1-5585)

209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys-----ArgLeu 226
 QY |||||
 Db |||||

104 TATCGCATCAACGGTAGTTATTCTGATCAGGCAATCGTGATACGCCGCGATGGACGCTCTG 163
 Db |||||

227 ProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLys 246
 QY |||||
 Db |||||

164 CCGAATACC-----AACTATCGTAAATAATGTCAGGGGTGTATGGTTGGTTATAACTCC 217
 Db |||||

247 ArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSer 266
 QY |||||
 Db |||||

218 GGAACCATCGT-----TTGGCCCTCGCTTGATCGCTACAGACTCGCGCAAACT 271
 Db |||||

267 HisGluTyrAspAspCysHisAlaAspIleLeuTrpGlnLysSerLeuLeuLeuLysArg 286
 QY |||||
 Db |||||

272 TAC---TATGAGGATCCAGCGGAAGCTAT-----GAGGCA 304
 Db |||||

287 TyrLeuGlnLeuTyrProHisLeuLeuThrGluGlu-----AspValAsp 301
 QY |||||
 Db |||||

305 TTTAGTGTCAAAATACCTAACTTGAACGAGAGAAAGTTGGGTATCTATGACACAGAC 364
 Db |||||

302 TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAspAlaHisAla 321
 QY |||||
 Db |||||

365 GTGGACGGTGACTATCAAAA-----AAAAATTCATTTCCGACGCG 403
 Db |||||

322 HisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrp 341
 QY |||||
 Db |||||

404 TATGACGACACCATCCAGCGCCCAATTTGCCAAC-----GAAGTAAAAACGACACAG 454
 Db |||||

342 LysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis 361
 QY |||||
 Db |||||

455 CTGTTCCAGTCGGATGATTCAGGCTCTGACCGTTTCAT---AACAGACTGAC---ACC 508
 Db |||||

362 HisAspGlnLysAlaGlyAspAlaVal-----GluAsnPhePheAsnAsnGlnThrGln 379
 QY |||||
 Db |||||

509 CATGATAAGCAATACACTCAGCGGGTCACATTCGAGAGTCACATTTTCGCTCGCTGCTAAT 568
 Db |||||

380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
 QY |||||
 Db |||||

569 AATGAATGTTTACC-----AsnTrp-----AspAsnPheThrLeu 444
 QY |||||
 Db |||||

701 CAAGTACAGTCTCATTTCGCACAAATGACATGGCGATTCGCCGATCACTGACATGG 760
 Db |||||

445 GluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle 464
 QY |||||
 Db |||||

761 ACAATGGAGTTCCG-----CAATACTGGCTTTCTTCAAGTTGACGCGTGGT----- 808
 Db |||||

465 AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla 484
 QY |||||
 Db |||||

809 GACGAGTATCATATACCGCAGCAATTAAGCATACCTCTCTTTCGACGAGAGTCTGGCG 868
 Db |||||

485 -----ArgSerPheAlaLeuSerGlyAsnTrpTyrPheTrpGlnHisLysLeuSer 502
 QY |||||
 Db |||||

869 AGTGATCAGGAATGGTAACATCTACAGCGCTGCTGCTATTCAGGTTTCGATAACTGGAG 928
 Db |||||

503 LeuThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuTyrAlaHis 520
 QY |||||
 Db |||||

929 TTACGGCTGCGTTTCGGCAAGCTACGTATTTCACACTCTCCAGCTTTTATGTCAG 988
 Db |||||

521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
 QY |||||
 Db |||||

989 ACATCTGCGGGCGGCGAGTGTACATAC-----GGAATCTCTATCTAAGGCTGAACAC 1042
 Db |||||

541 SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAla 560
 QY |||||
 Db |||||

1043 TCCAATAACTTTGAATTAGTGCACGATATAATGTAATCAGTGGCTGATTGACACGCGCA 1102
 Db |||||

561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
 QY |||||
 Db |||||

1103 GTTACTACTCAGACCTAAAGATTATAT---GCAAGTCTGATCTGTGATGCGAGTATA 1159
 Db |||||

581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr-----AsnGlnSer 598
 QY |||||
 Db |||||

1160 GTTTGCAATGGTAAACACCACTCTCCGTAGTAGCTACTATTATTATGACAAATATTGAT 1219
 Db |||||

599 GlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArg 618
 QY |||||
 Db |||||

1220 CGGGCAAAACATGGGGGACTGGAA-----ATAAGCGGGAATATAAT 1261
 Db |||||

619 IleGlyValSerGlyAspTyrValArgGlyArgLeu----- 630
 QY |||||

1262 GGTCTGGTTTTCTCGCCATATATCAGTGGCAATTTAATTCGTGCGCAATATGAACCTCA 1321
 Db |||||

631 -----LysAsnLeuProSerLeuProGlyArg----- 639
 QY |||||

1322 ACATTAATAAACAACTAATAACAGGAGAACCCAGCGATAAACGAGCTATAGGGCTGAAACAT 1381
 Db |||||

640 -----GluAspAlaTyrGlyAsnArgProPheIle 649
 QY |||||

1382 ACTCTTGTGATGGTCCAGCCAACTAATCTCTGATGTTTATTTATTCGTGCTCTAGT 1441
 Db |||||

650 AlaGlnAspAspGlnAsnAla-----ProArgValProAla---AlaArgLeuGlyPhe 666
 QY |||||

1442 GCAAAAGATGACAGTAACTGACCGGAAACAAATGTTTCGGGCTGGGCACTCTCAACTTT 1501
 Db |||||

667 HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal--- 685
 QY |||||

1502 GCATTAATACAGAAATTC-----GGTAACGAGGATCAGTACCGGATTAAC 1546
 Db |||||

686 PheAlaGlnAsnLysLeuAla-----ArgTyrGluThr-----ArgThrProGly 700
 QY |||||

1547 CTAGCACTCAATTAACCTGACAGCAAACTCCGTCACGACCATGAATATTCCTGCA 1606
 Db |||||

701 -HisHisMetLeuAsnLeu-----GlyAlaAsnTyrArgArgAsnThrArgTyr 716
 QY |||||

1607 GCAGGTTTAAATGCACTATAGTGTGTTGATGGAATTTCTGATGACAAACAAATCGTTA 1666
 Db |||||

716 rGly 717
 QY |||||

1667 TGGG 1670
 Db |||||

RESULT 28
 US-10-085-959-58
 ; Sequence 58, Application US/10085959
 ; Publication No. US20030165870A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Blattner, Frederick R.
 ; APPLICANT: Welch, Rodney A.
 ; APPLICANT: Burland, Valerie D.
 ; TITLE OF INVENTION: No. US20030165870A1el Sequence of E. Coli CFT073
 ; FILE REFERENCE: 960296.97648
 ; CURRENT APPLICATION NUMBER: US/10/085,959
 ; PRIOR FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: 60/242,412
 ; NUMBER OF SEQ ID NOS: 255
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 58
 ; LENGTH: 8249
 ; TYPE: DNA

ORGANISM: Escherichia coli
US-10-085-959-58

Alignment Scores:

Pred. No.: 3,646-06 Length: 8249
Score: 160.50 Matches: 129
Percent Similarity: 36.25% Conservatives: 82
Best Local Similarity: 22.16% Mismatches: 239
Query Match: 3.98% Indels: 132
DB: 12 Gaps: 33

US-09-936-377-2 (1-758) x US-10-085-959-58 (1-8249)

```

Qy 209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys-----ArgLeu 226
Db 116 TATGCATCAACGGTAGTTATCTGTATCAGGCAATCGTACCGCGGATGGACGCTG 175
Qy 227 ProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLys 246
Db 176 CCGAATACC-----AACTATCGTAACAATAGTCAGGGTGATGGTTGGTTAACTCC 229
Qy 247 ArgPheTyrArgThrTyrSerAspArgAspGlnTyrGlyLeuProAlaHisSer 266
Db 230 GGAACCATCGT-----TTTGGCTCTCGCTTACGTACAGACTCGCGACCAAACT 283
Qy 267 HisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuLeuAsnLysArg 286
Db 284 TAC---TATGAGGATCCAGACGGAAGCTAT-----GAGGCA 316
Qy 287 TyrLeuGlnLeuTyrProHisLeuLeuThrGluGlu-----AspValAsp 301
Db 317 TTTAGTGCAAAATACCTAACTTGAACGAGAGAAAGTTGGGTATCTATGACACAGAC 376
Qy 302 TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAla 321
Db 377 GTGACGCGTACTATCAAAA-----AAAATTCATTTCCGACGCG 415
Qy 322 HisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrp 341
Db 416 TATGACGACAGCATCCAGCGCAATTTGCCAAC-----GAAGTAAACAAACACACAG 466
Qy 342 LysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis 361
Db 467 CCTGTTCCAGTCGATGATTCAGGCTCTGACGCTTCAT---AACAGACTGAC---ACC 520
Qy 362 HisAspGluLysAlaGlyAspAlaVal-----GluAsnPhePheAsnAsnGlnThrGln 379
Db 521 CATGATAAGCAATACTACACTCAGCGCGTCACATTGCAGAGTCACCTTTTCGCTGCTGCTAAT 580
Qy 380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
Db 581 AATGAACTTGTACC-----AsnTrp-----GGTGCA 601
Qy 400 GlnTyr-----LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla 415
Db 602 CAGTACAAACAGACAGGTCAGCAAGGTCGGTGCATGACCTCAAGCAATCTCTG 661
Qy 416 ValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlu 435
Db 662 ACCGCGC-----TTCAATTAATAAGGAAACACGAACTCGCTCTCTATTATGAGTCAGAG 712
Qy 436 GlnAla-----AsnTrp-----AspAsnPheThrLeu 444
Db 713 CAAGTACAGCTCTACTATTGCACAAAATAGTCGCGATTCGCGCATCACTGGACATGG 772
Qy 445 GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle 464
Db 773 ACAATGGAGTTGCG-----CAATACGGCTTCTTCAAAAGTTGACGCGTGGT----- 820
Qy 465 AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla 484
Db 821 GACGAGGATCATATACCGCGAGCATTTATAGCGATACCTCTCTTGGCAGAGAGTCTGCG 880

```

```

Qy 485 -----ArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer 502
Db 881 AGTGATCACAAATGGTAACATCTACAAGCGCTCGCTATTACAGGTTTCGATAACTTGGAG 940
Qy 503 LeuThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuTyrAlaHis 520
Db 941 TTACGCGCTGGTTCGCGAAGGCTAGTATTTCCACACTCTCCAGCTTTTATGCGAG 1000
Qy 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
Db 1001 ACATCTCGGCGCGCAGTGTCAATAC-----GGAATCTCTGATCTTAAGCGCTGAACAC 1054
Qy 541 SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAla 560
Db 1055 TCCAAATACCTTGAATAGGTGCGCATATAATGTAATCAGTGGCTGATGACAGCGCA 1114
Qy 561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
Db 1115 GTTTACTACTCAGAAAGCTAAAGATTATATT---GCAAGTCTGATCTGTGATGGCAGTATA 1171
Qy 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr-----AsnGlnSer 598
Db 1172 GTTTGCAATGTAAACCACTCTCCGCTAGTAGTACTATTATTATGACAATATTGAT 1231
Qy 599 GlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArg 618
Db 1232 CGGCAAAACATCGGCGACTGGA-----ATAGCGCGGAATATATAAT 1273
Qy 619 IleGlyValSerGlyAspTyrValArgGlyArgLeu----- 630
Db 1274 GGCTGGGTTTCTCGCCATATATCAGTGGCAATTTAATTCGTCGCAATATGAAACTTCA 1333
Qy 631 -----LysAsnLeuProSerLeuProGlyArg----- 639
Db 1334 ACATTAAAAACAATAATACAGAGAACCCAGCGATAAACCGACGTATAGGCTGAAACAT 1393
Qy 640 -----GluAspAlaTyrGlyAsnArgProPheIle 649
Db 1394 ACTCTGTGATGGGTGAGCCCAACATATCTCTGATGTTTTTATTCGTGCTGCTCTAGT 1453
Qy 650 AlaGlnAspAspGlnAsnAla-----ProArgValProAla---AlaArgLeuGlyPhe 666
Db 1454 GCAAAAGATGACAGTAACGTTACCGAACAATGTTCCGGGTGCGGCACCTCAACTTT 1513
Qy 667 HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal--- 685
Db 1514 GCAGTAAATACAGAAATTC-----GGTAAACGAGGATCAGTACCGGATTAAC 1558
Qy 686 PheAlaGlnAsnLysLeuAla-----ArgTyrGluThr-----ArgThrProGly 700
Db 1559 CTAGACCTCAATAACCTGACAGAACCGTACCGTACGACATGAAACTATTCTCTGCA 1618
Qy 701 -HisHisMetLeuAsnLeu-----GlyAlaAsnTyrArgArgAsnThrArgTyr 716
Db 1619 GCAGGTTTAAATGCGCTATAGGTTTCTGTATGGAATTTCTGTATGACAAAAACAATCGTTA 1678
Qy 716 rGly 717
Db 1679 TGA 1682

```

RESULT 29

US-10-329-960-1/c

; Sequence 1, Application US/10329960

; Publication No. US20030099277A1

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Pr

; TITLE OF INVENTION: Thereof, and Uses Thereof

; FILE REFERENCE: PBI86P1

; CURRENT APPLICATION NUMBER: US/10/329,960

; CURRENT FILING DATE: 2003-01-02

; PRIOR APPLICATION NUMBER: US 09/643,990

; PRIOR FILING DATE: 2000-08-23

```

1, PRIOR APPLICATION NUMBER: US 08/487,429
2,
3, PRIOR FILING DATE: 1995-06-07
4, PRIOR APPLICATION NUMBER: US 08/426,787
5, PRIOR FILING DATE: 1995-04-21
6, NUMBER OF SEQ ID NOS: 1
7, SOFTWARE: PatentIn version 3.1
8, SEQ ID NO 1
9, LENGTH: 1830121
10, TYPE: DNA
11, ORGANISM: Haemophilus influenzae
12,
13, FEATURE:
14, NAME/KEY: misc feature
15, LOCATION: (4747)..(4747)
16, OTHER INFORMATION: n equals a, t, g or c
17, FEATURE:
18, NAME/KEY: misc feature
19, LOCATION: (9921)..(9921)
20, OTHER INFORMATION: n equals a, t, g or c
21, FEATURE:
22, NAME/KEY: misc feature
23, LOCATION: (10150)..(10150)
24, OTHER INFORMATION: n equals a, t, g or c
25, FEATURE:
26, NAME/KEY: misc feature
27, LOCATION: (29298)..(29298)
28, OTHER INFORMATION: n equals a, t, g or c
29, FEATURE:
30, NAME/KEY: misc feature
31, LOCATION: (36543)..(36543)
32, OTHER INFORMATION: n equals a, t, g or c
33, FEATURE:
34, NAME/KEY: misc feature
35, LOCATION: (36551)..(36551)
36, OTHER INFORMATION: n equals a, t, g or c
37, FEATURE:
38, NAME/KEY: misc feature
39, LOCATION: (36636)..(36636)
40, OTHER INFORMATION: n equals a, t, g or c
41, FEATURE:
42, NAME/KEY: misc feature
43, LOCATION: (40808)..(40810)
44, OTHER INFORMATION: n equals a, t, g or c
45, FEATURE:
46, NAME/KEY: misc feature
47, LOCATION: (44416)..(44416)
48, OTHER INFORMATION: n equals a, t, g or c
49, FEATURE:
50, NAME/KEY: misc feature
51, LOCATION: (44905)..(44905)
52, OTHER INFORMATION: n equals a, t, g or c
53, FEATURE:
54, NAME/KEY: misc feature
55, LOCATION: (44975)..(44975)
56, OTHER INFORMATION: n equals a, t, g or c
57, FEATURE:
58, NAME/KEY: misc feature
59, LOCATION: (45593)..(45593)
60, OTHER INFORMATION: n equals a, t, g or c
61, FEATURE:
62, NAME/KEY: misc feature
63, LOCATION: (45732)..(45732)
64, OTHER INFORMATION: n equals a, t, g or c
65, FEATURE:
66, NAME/KEY: misc feature
67, LOCATION: (47036)..(47036)
68, OTHER INFORMATION: n equals a, t, g or c
69, FEATURE:
70, NAME/KEY: misc feature
71, LOCATION: (51334)..(51334)
72, OTHER INFORMATION: n equals a, t, g or c
73, FEATURE:
74, NAME/KEY: misc feature
75, LOCATION: (51602)..(51602)
76,
77, OTHER INFORMATION: n equals a, t, g or c
78, FEATURE:
79, NAME/KEY: misc feature
80, LOCATION: (51786)..(51786)
81, OTHER INFORMATION: n equals a, t, g or c
82, FEATURE:
83, NAME/KEY: misc feature
84, LOCATION: (51805)..(51805)
85, OTHER INFORMATION: n equals a, t, g or c
86, FEATURE:
87, NAME/KEY: misc feature
88, LOCATION: (55369)..(55369)
89, OTHER INFORMATION: n equals a, t, g or c
90, FEATURE:
91, NAME/KEY: misc feature
92, LOCATION: (65309)..(65309)
93, OTHER INFORMATION: n equals a, t, g or c
94, FEATURE:
95, NAME/KEY: misc feature
96, LOCATION: (65313)..(65313)
97, OTHER INFORMATION: n equals a, t, g or c
98, FEATURE:
99, NAME/KEY: misc feature
100, LOCATION: (80024)..(80024)
101, OTHER INFORMATION: n equals a, t, g or c
102, FEATURE:
103, NAME/KEY: misc feature
104, LOCATION: (100091)..(100091)
105, OTHER INFORMATION: n equals a, t, g or c
106, FEATURE:
107, NAME/KEY: misc feature
108, LOCATION: (102696)..(102696)
109, OTHER INFORMATION: n equals a, t, g or c
110, FEATURE:
111, NAME/KEY: misc feature
112, LOCATION: (105121)..(105121)
113, OTHER INFORMATION: n equals a, t, g or c
114, FEATURE:
115, NAME/KEY: misc feature
116, LOCATION: (107248)..(107248)
117, OTHER INFORMATION: n equals a, t, g or c
118, FEATURE:
119, NAME/KEY: misc feature
120, LOCATION: (117136)..(117136)
121, OTHER INFORMATION: n equals a, t, g or c
122, FEATURE:
123, NAME/KEY: misc feature
124, LOCATION: (119750)..(119750)
125, OTHER INFORMATION: n equals a, t, g or c
126, FEATURE:
127, NAME/KEY: misc feature
128, LOCATION: (119924)..(119924)
129, OTHER INFORMATION: n equals a, t, g or c
130, FEATURE:
131, NAME/KEY: misc feature
132, LOCATION: (120038)..(120038)
133, OTHER INFORMATION: n equals a, t, g or c
134, FEATURE:
135, NAME/KEY: misc feature
136, LOCATION: (121344)..(121344)
137, OTHER INFORMATION: n equals a, t, g or c
138, FEATURE:
139, NAME/KEY: misc feature
140, LOCATION: (122167)..(122167)
141, OTHER INFORMATION: n equals a, t, g or c
142, FEATURE:
143, NAME/KEY: misc feature
144, LOCATION: (122336)..(122336)
145, OTHER INFORMATION: n equals a, t, g or c
146, FEATURE:
147, NAME/KEY: misc feature
148, LOCATION: (131340)..(131340)
149, OTHER INFORMATION: n equals a, t, g or c
150, FEATURE:
151, NAME/KEY: misc feature
152, LOCATION: (51602)..(51602)
153,
154, OTHER INFORMATION: n equals a, t, g or c
155, FEATURE:
156, NAME/KEY: misc feature
157, LOCATION: (51786)..(51786)
158, OTHER INFORMATION: n equals a, t, g or c
159, FEATURE:
160, NAME/KEY: misc feature
161, LOCATION: (51805)..(51805)
162, OTHER INFORMATION: n equals a, t, g or c
163, FEATURE:
164, NAME/KEY: misc feature
165, LOCATION: (55369)..(55369)
166, OTHER INFORMATION: n equals a, t, g or c
167, FEATURE:
168, NAME/KEY: misc feature
169, LOCATION: (65309)..(65309)
170, OTHER INFORMATION: n equals a, t, g or c
171, FEATURE:
172, NAME/KEY: misc feature
173, LOCATION: (65313)..(65313)
174, OTHER INFORMATION: n equals a, t, g or c
175, FEATURE:
176, NAME/KEY: misc feature
177, LOCATION: (80024)..(80024)
178, OTHER INFORMATION: n equals a, t, g or c
179, FEATURE:
180, NAME/KEY: misc feature
181, LOCATION: (100091)..(100091)
182, OTHER INFORMATION: n equals a, t, g or c
183, FEATURE:
184, NAME/KEY: misc feature
185, LOCATION: (102696)..(102696)
186, OTHER INFORMATION: n equals a, t, g or c
187, FEATURE:
188, NAME/KEY: misc feature
189, LOCATION: (105121)..(105121)
190, OTHER INFORMATION: n equals a, t, g or c
191, FEATURE:
192, NAME/KEY: misc feature
193, LOCATION: (107248)..(107248)
194, OTHER INFORMATION: n equals a, t, g or c
195, FEATURE:
196, NAME/KEY: misc feature
197, LOCATION: (117136)..(117136)
198, OTHER INFORMATION: n equals a, t, g or c
199, FEATURE:
200, NAME/KEY: misc feature
201, LOCATION: (119750)..(119750)
202, OTHER INFORMATION: n equals a, t, g or c
203, FEATURE:
204, NAME/KEY: misc feature
205, LOCATION: (119924)..(119924)
206, OTHER INFORMATION: n equals a, t, g or c
207, FEATURE:
208, NAME/KEY: misc feature
209, LOCATION: (120038)..(120038)
210, OTHER INFORMATION: n equals a, t, g or c
211, FEATURE:
212, NAME/KEY: misc feature
213, LOCATION: (121344)..(121344)
214, OTHER INFORMATION: n equals a, t, g or c
215, FEATURE:
216, NAME/KEY: misc feature
217, LOCATION: (122167)..(122167)
218, OTHER INFORMATION: n equals a, t, g or c
219, FEATURE:
220, NAME/KEY: misc feature
221, LOCATION: (122336)..(122336)
222, OTHER INFORMATION: n equals a, t, g or c
223, FEATURE:
224, NAME/KEY: misc feature
225, LOCATION: (131340)..(131340)
226, OTHER INFORMATION: n equals a, t, g or c
227, FEATURE:
228, NAME/KEY: misc feature
229, LOCATION: (51602)..(51602)
230,
231, OTHER INFORMATION: n equals a, t, g or c
232, FEATURE:
233, NAME/KEY: misc feature
234, LOCATION: (51786)..(51786)
235, OTHER INFORMATION: n equals a, t, g or c
236, FEATURE:
237, NAME/KEY: misc feature
238, LOCATION: (51805)..(51805)
239, OTHER INFORMATION: n equals a, t, g or c
240, FEATURE:
241, NAME/KEY: misc feature
242, LOCATION: (55369)..(55369)
243, OTHER INFORMATION: n equals a, t, g or c
244, FEATURE:
245, NAME/KEY: misc feature
246, LOCATION: (65309)..(65309)
247, OTHER INFORMATION: n equals a, t, g or c
248, FEATURE:
249, NAME/KEY: misc feature
250, LOCATION: (65313)..(65313)
251, OTHER INFORMATION: n equals a, t, g or c
252, FEATURE:
253, NAME/KEY: misc feature
254, LOCATION: (80024)..(80024)
255, OTHER INFORMATION: n equals a, t, g or c
256, FEATURE:
257, NAME/KEY: misc feature
258, LOCATION: (100091)..(100091)
259, OTHER INFORMATION: n equals a, t, g or c
260, FEATURE:
261, NAME/KEY: misc feature
262, LOCATION: (102696)..(102696)
263, OTHER INFORMATION: n equals a, t, g or c
264, FEATURE:
265, NAME/KEY: misc feature
266, LOCATION: (105121)..(105121)
267, OTHER INFORMATION: n equals a, t, g or c
268, FEATURE:
269, NAME/KEY: misc feature
270, LOCATION: (107248)..(107248)
271, OTHER INFORMATION: n equals a, t, g or c
272, FEATURE:
273, NAME/KEY: misc feature
274, LOCATION: (117136)..(117136)
275, OTHER INFORMATION: n equals a, t, g or c
276, FEATURE:
277, NAME/KEY: misc feature
278, LOCATION: (119750)..(119750)
279, OTHER INFORMATION: n equals a, t, g or c
280, FEATURE:
281, NAME/KEY: misc feature
282, LOCATION: (119924)..(119924)
283, OTHER INFORMATION: n equals a, t, g or c
284, FEATURE:
285, NAME/KEY: misc feature
286, LOCATION: (120038)..(120038)
287, OTHER INFORMATION: n equals a, t, g or c
288, FEATURE:
289, NAME/KEY: misc feature
290, LOCATION: (121344)..(121344)
291, OTHER INFORMATION: n equals a, t, g or c
292, FEATURE:
293, NAME/KEY: misc feature
294, LOCATION: (122167)..(122167)
295, OTHER INFORMATION: n equals a, t, g or c
296, FEATURE:
297, NAME/KEY: misc feature
298, LOCATION: (12233
```

```

FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.:
Score: 0.0119
Percent Similarity: 160.50
Best Local Similarity: 32.42%
Query Match: 19.35%
Indels: 371
Gaps: 50

US-09-936-377-2 (1-758) x US-10-329-960-1 (1-1830121)

QY 5 ThrLeuLysProfile-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db 1059601 ACTAAAAACCCCTATTTTCGCTAAGTATATTTCTTGCTTTAATTCATGCTATGTA 1059542
QY 21 LeuAlaGlnAlaHis-----GluThrGlnSerValGly-----32
Db 1059541 AAAGCAGAACTCAAAGTATAAAGATACAAAGAGCTATATCATCTGAAGTGGACACT 1059482
QY 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
Db 1059481 CAAAGTACAGAAGATTTCAGAAATTTAGAACTATCTCAGTCACTGCAGAAAAGTAAGAGAT 1059422
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 1059421 CGTAAAGATAATCAAGTA-----ACTGGACTTGGCAAAATATCAAACTAGT 1059374
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 1059373 GAAAGTATCAGCCGAGAACCAAGTATTAATATTTTCGTGATCTAAACGCTATGATCCAGGG 1059314

```


QY 389 oileglyArgLeuLysGlySerTrpGlyValGlnTrpLeuGlyGlnLysSerSerAlaLe 409
 Db 1058386 -----GGTATTGATATATTTTACGAAAAATAAGAACAAACAGC 1058352
 QY 409 uSerAlaThrSerGluAlaVal-----LysGlnProMetLeuLeuAsnLys 425
 Db 1058351 GGGCATCATTTGCAAGAGCGTGTAAAGTCTAATCAACAAAACATCATCTTGCACAGTTA 1058292
 QY 425 sValGlnHis-----428
 Db 1058291 TAIGCGACATACGATTCAGCTCTTATCTCTAATCAAGTAAGATTGGCGCCCAACACT 1058232
 QY 429 -----TyrSerPhePheGlyValGlu-----435
 Db 1058231 TGATAAACCTTATTCATCTATCTGTTCTGATGAAATGTTTATAAGAAAAACATAATAT 1058172
 QY 436 -----GlnAlaAsnTrp-----439
 Db 1058171 GTTGCAATTGAATTTAGAGAAAAAATTCACAAAATTTGGCTTACTCATCAAAATTCCTTT 1058112
 QY 440 -----AspAsnPheThr-----443
 Db 1058111 CAATCTTGTTTGTAGACTTTTCTTCCGACCTTCAGCATAAAGATTATTTRAATCGACG 1058052
 QY 443 -----443
 Db 1058051 TGTATCGCTACGGCAGATAGTATTTTCAGATAAACTGCTTAAACAAAGAAATGGTTT 1057992
 QY 444 -----LeuGluGlyValArgVa 450
 Db 1057991 ACCGGAATACCTTACTTATACCCAAAACCAAGACCATATTTTGCAGGAGAAATCATTCG 1057932
 QY 450 lGluLysGlnLysAlaSerIleArgTyrAspLysAla-----LeuIleAspArgGl 467
 Db 1057931 TAATTATCAGGTAGTTCTCTCTAATTACAGAGACTGTAAAGTGGCTTAAATAAGCGAA 1057872
 QY 467 uAsnTyrTyr-----LysGlnProLeuProAspLeuGly-- 478
 Db 1057871 AAATTATTATTCGCACACGCAATATATGCGATTAGGAAATACGTTGATTAGGTTT 1057812
 QY 479 -----AlaHisArgGlnThrAla-----484
 Db 1057811 AGGTATTCGTATGACGTATCTCGTACAAAAGCTTAATGAATCACTATTAGTTGGTAA 1057752
 QY 485 ----ArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLe 503
 Db 1057751 ATTTAAAAATTTCTCTTGGAACTACTGTTATGTTCATAAAACCAACGGAATGGCTTGATCT 1057692
 QY 503 uThr-----AlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGl 521
 Db 1057691 TTCTTATCGCCTTTCTACTCGATTAGAAATCTTAGTTTTCGTAATGTAT-----GG 1057638
 QY 521 YLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgse 541
 Db 1057637 TTGGCGGTATGGTGGCAAGATGACGAGGTTTATGTAGGTAAATTTTAAGCCTGAACATC 1057578
 QY 541 rAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLe 561
 Db 1057577 TCGTAAACCAAGAGTTTGGTCTCGCTCTAAAGAGGGGATTTTGGTAATATTGAGATCACTCA 1057518
 QY 561 uTyrArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsn-----As 577
 Db 1057517 TTTTAGTAATGCTTATCGAAATCTTATCGCTTTGCTGGAACACTTAGTAAAAATGGAAC 1057458
 QY 577 pGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsn-- 596
 Db 1057457 TGGAAAGGCAATTTATGATATCATTAATGCACAAAATGCAAAATTTAGTTGGTGTAAATAT 1057398
 QY 597 -GlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe-----611
 Db 1057397 AACTGCGCAATTAGATTTTATGTTTATGTTTATGAAACGATTATTCCTACGTTGGTATGCAAC 1057338
 QY 611 -----611

Db 1057337 ATTTCCTTATACCGAGTAAAGTTAAAGATCAAAAAATCAATGCTGGTTTACCTCCGT 1057278
 QY 612 -----LysProThrProArgTyrArgIleGlyValSerGl 623
 Db 1057277 AAGCAGTTATTTTATTTGATGCCATTCAGCCAGC---CGTTATATCATTTGGTTTAGGCTA 1057221
 QY 623 yAspTyrVal-----ArgGlyArgLeuLysAs 632
 Db 1057220 TGATCATATAAGTAATACTTTGGGAGTAATGCTACCTTTTACCCATCAAAAGCAAAATC 1057161
 QY 632 nLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAlaGlnAs 652
 Db 1057160 TCAAAATGAATTTGCTAGGAAAACGT---GCATTAGGTAAACAT---TCAAGGGA 1057113
 QY 652 pAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLe 672
 Db 1057112 TGTAAATCAACAAAGAAATCTTACTCGGCA-----TGCATATC----- 1057073
 QY 672 uThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAl 692
 Db 1057072 -----TTAGATGTATCGGGTTATTACATCGCGAATAAA----- 1057040
 QY 692 aArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAla----- 708
 Db 1057039 -----AATATTATGCTTCGATTAGGATATATAATTATT 1057005
 QY 709 -AsnTyrArg-----ArgAsnThrArgTyrGlyGluTrpAsnTr 721
 Db 1057004 CAACTATCGTTATGTTTACTTGGGAACGGTTCGCTCAACAGCACAAAGTTCGGTCAATCA 1056945
 QY 721 pTyrValLysAlaAspAsnLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSe 741
 Db 1056944 ACAT-----CAAAATGTTGGTAACTATCTCGTACGACGACATCA----- 1056905
 QY 741 rAspThrProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
 Db 1056904 -----GGACGAAATATACCTTAAACATTAGAAATGAAATTC 1056869
 RESULT 30
 US-10-238-075-720
 ; Sequence 720, Application US/10238075
 ; Publication No. US20030148324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isola
 ; FILE OF INVENTION: E.coli, and biological uses of these polynucleotides and of the
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238, 075
 ; CURRENT FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 0003145
 ; PRIOR FILING DATE: 2000-03-10
 ; NUMBER OF SEQ ID NOS: 1576
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 720
 ; LENGTH: 1620
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 US-10-238-075-720
 Alignment Scores:
 Pred. No.: 3.6e-07 Length: 1620
 Score: 160.00 Matches: 122
 Percent Similarity: 36.17% Conservatives: 78
 Best Local Similarity: 22.06% Mismatches: 229
 Query Match: 3.96% Indels: 124
 DB: 12 Gaps: 31
 US-09-936-377-2 (1-758) x US-10-238-075-720 (1-1620)
 QY 209 TyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLys-----ArgLeu 226
 Db 79 TATCGCATCAACGGTAGTTATTTCTGATCAGGCAATCTGATAGCCGGATGGACGCTG 138

QY 227 ProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLys 246
 DB 139 CGGAATACC-----AACTATCGTAACAATAGTCAGGGTGTATGGTTGGTTGATATAACTCC 192
 QY 247 ArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSer 266
 DB 193 GGAACCATCGT-----TTTGCCCTCTCGCTTGCATCGCTACAGATCGCGACGCAAACT 246
 QY 267 HisGluTyrAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLysArg 286
 DB 247 TAC---TATGAGGATCCAGACGGAACCTAT-----GAGGCA 279
 QY 287 TyrLeuGlnLeuTyrProHisLeuLeuThrGluGlu-----AspValAsp 301
 DB 280 TTTAGTGTCAAAATACCTAACTTGAACGAGAGAAAGTTGGGTATTTCTATGACACAGAC 339
 QY 302 TyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAla 321
 DB 340 GTGGACGGTGACTATCTAAAA-----AAAATTCAATTCGAGCGC 378
 QY 322 HisAsnGlyLysProTrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrp 341
 DB 379 TATGACGACACCATCCAGCGCAATTTGCCAAC-----GAAGTAAAAACGACACAG 429
 QY 342 LysGlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis 361
 DB 430 CCTGTTCCTCCAGTCCGATGATTCAGGCTCTGACCGTTTCAT---AACAGACTGAC---ACC 483
 QY 362 HisAspGluLysAlaGlyAspAlaVal-----GluAsnPhePheAsnAsnGlnThrGln 379
 DB 484 CATGATAGCAATACACTCAGCGGTGCACATTCGAGAGTCACATTTTCGCTGCCTGCTAAT 543
 QY 380 AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyVal 399
 DB 544 AATGAACCTGTGTACC-----GGTGCA 564
 QY 400 GlnTyr-----LeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla 415
 DB 565 CAGTACAAACACAGACAGGGTCAAGCAAGTTCGGTGGCATGACCTCAAGCAATCTCTG 624
 QY 416 ValLysGlnProMetLeuLeuAsnLysValGlnHisTyrSerPhePheGlyValGlu 435
 DB 625 ACCGGC-----TTCAATATATAGAAACACAGAACTCGCTCTTATATAGTACAGAG 675
 QY 436 GlnAla-----AsnTrp-----AspAsnPheThrLeu 444
 DB 676 CAAAGTACAGTCTCACTATTTCGCACAAATAGTGGCGATTGCGCGATCACTGGACATGG 735
 QY 445 GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle 464
 DB 736 ACAATGGGAGTTCGC-----CAATCTGGCTTTCTTCAAGTTGACGCGTGGT----- 783
 QY 465 AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla 484
 DB 784 GACGAGTATCATATACCGCAGGATTAAGCATACCTCTCTTGCAGAGATCTGCG 843
 QY 485 -----ArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSer 502
 DB 844 AGTGATCACGAAATGGTAACATCTACAAAGCGCTGCGCTATTTCAGGTTTCGATACTGGAG 903
 QY 503 LeuThrAlaSerHisGlnGluArg-----LeuProSerThrGlnGluLeuTyrAlaHis 520
 DB 904 TTACGCGCTGGTTCGCGCAAGGCTAGTATTTCCACACTCTCCAGCTTTTATGAG 963
 QY 521 GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArg 540
 DB 964 ACATCTCGGCGCGCAGTGTACATAC-----GGAATCTCTGATCTTAAGCTTGAACAC 1017
 QY 541 SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAla 560
 DB 1018 TCCAAATAACTTTGAATTAGGTGCAGATATATGTAATCATGTCGCTGATTGACAGCGCA 1077

561 LeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGly 580
 1078 GTTTACTACTCAGAACGTAAGATATATTT---GCAAGTCTGATCTGATGCGCATATA 1134
 581 ProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr-----AsnGlnSer 598
 1135 GTTTCGAATGGTAACACCAACTCTCCCGTAGTAGTACTATTATTATGACAATATTGAT 1194
 599 GlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgTyrArg 618
 1195 CGGGCAAAACATGGGGACTGGAA-----ATAAGCGCGGAATATAAT 1236
 619 IleGlyValSerGlyAspTyrValArgGlyArgLeu----- 630
 1237 GGTGGGTGTTTCTCGCATATATCACTGGCAATTAATTCGTCGCGCAATATGAACCTCA 1296
 631 -----LysAsnLeuProSerLeuProGlyArg----- 639
 1297 ACATTAATAACAATAATACAGGAGAACCGAGATTAACGGCTATAGGGCTGAAACAT 1356
 640 -----GluAspAlaTyrGlyAsnArgProPheIle 649
 1357 ACTCTTGTGATGGTCCAGCCACATAATCTCTGTATGTTTATTCTGCTGCTCTAGT 1416
 650 AlaGlnAspAspGlnAsnAla-----ProArgValProAla---AlaArgLeuGlyPhe 666
 1417 GCAAAAGATGACAGTAAACGGTACCGAAACAAATGTTCCGGGCTGGGCCACTCTCACTTT 1476
 667 HisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal--- 685
 1477 GCAGTAATAACAGAAATTC-----GGTAACGAGGATCAGTACCGGATTAAC 1521
 686 PheAlaGlnAsnLysLeuAla-----ArgTyrGluThr 696
 1522 CTAGCACTCAATAACCTGACAGACAAACGCTACCGGTACA 1560

RESULT 31
 US-10-114-170-57/c
 Sequence 57, Application US/10114170
 Publication No. US20030023075A1
 GENERAL INFORMATION:
 APPLICANT: Blattner, Frederick R.
 Burland, Valerie
 Perna, Nicole T.
 Plunkett, Guy
 Welch, Rod
 TITLE OF INVENTION: NO. US20030023075A1el Sequences of E. coli O157
 NUMBER OF SEQUENCES: 265
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Charles & Brady
 STREET: 1 South Pinckney Street
 CITY: Madison
 STATE: WI
 ZIP: 53701-2113
 COUNTRY: US
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 8.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/114,170
 FILING DATE: 01-Apr-2002
 CLASSIFICATION: <unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/453,702
 FILING DATE: 03-DEC-1999
 APPLICATION NUMBER: 60/110,955
 FILING DATE: 04-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 960296.95017


```
QY 569 rIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspSerG1 589
Db 48480 GATTGCTCTTATCCATAAATGAT-----AACACCAATAGCTATGTAACACGCGAAA 48427
QY 589 uMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyL1 609
Db 48426 GGCCCGGTG-----CACCGTGTGGAATTGCGGCACA-----48393
QY 609 eTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyAr 629
Db 48392 -----TTGCCGCTGTGTCAGAGATGTCACGCTGCTCACTGAATTACCTGGACCG 48340
QY 629 gLeuLysAsnLeuProSerLeuProGlyArgGlyAlaAspAlaTyrGlyAsnArgProPheI1 649
Db 48339 A-----ACTGAACACACGTGATGGT-----48321
QY 649 eAlaGlnAspAspGlnAsnAlaPro-----ArgValProAlaAlaArgLeuGlyPheH1 667
Db 48320 -----GATAACAAAGGTGCGCGCTGAGTTATACCCCTGAACACATGTTGATGCGAA 48268
QY 667 sLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheA1 687
Db 48267 ACTGAAGTGCAGATCACCGAAGAGGTGGCATCATGCTGGT-----48225
QY 687 aGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuG1 707
Db 48224 -----GCCCGTTATCGCGGGAACACCA-----CGTTTCAC 48193
QY 707 yAlaAsnTyrArg-----ArgAsnThrArgTyrGlyGluTrpAsnTrpTy 722
Db 48192 CCAGAATTATTCGTCACAGCGCTGTACAGAGAAAGTGTATGATGAGAAAGAGAATA 48133
QY 722 rValLysAla-----725
Db 48132 CCTGAAGCCTGGACGGTGGATGCGAGTCTCTCGTGAAGATGACGGATGCCCTGAC 48073
QY 726 -----AspLeuLeuLeuAsnGln-----SerValTyrAl 735
Db 48072 GCTGAATGCTGCGGTGAATACCTGCTCAACAGGATTACAGTACGCTGTACAG 48013
QY 735 a-----HisSerPheLeuSerAspThrProGlnMetGlyArgSerPheThrG1 752
Db 48012 TGCCGGTAAGAGTACGCTGTATGCGGTGATTACTTCAGACGGATCATCAACACAGG 47953
QY 752 yGlyVal 754
Db 47952 ATATGTG 47946
```

RESULT 32

```
US-10-098-808-1
; Sequence 1, Application US/10098808
; Publication No. US20030007981A1
; GENERAL INFORMATION:
; APPLICANT: Potter, Andrew A.
; APPLICANT: Rioux, Clement
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
; TITLE OF INVENTION: TRANSFERRIN-BINDING PROTEINS
; FILE REFERENCE: 9000-0049.20
; CURRENT APPLICATION NUMBER: US/10/098,808
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/405,728
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/267,749
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6376
; TYPE: DNA
; ORGANISM: Haemophilus somnus
US-10-098-808-1
```

Alignment Scores:

```
Pred. No.: 7.7e-06 Length: 6376
Score: 156.00 Matches: 195
Percent Similarity: 30.43% Conservative: 120
Best Local Similarity: 18.84% Mismatches: 366
Query Match: 3.87% Indels: 354
DB: 14 Gaps: 50

US-09-936-377-2 (1-758) x US-10-098-808-1 (1-6376)
QY 6 LeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeu-----21
Db 2915 CTTAAGCTGATACATTTGGCTGTCAGCAGCATTTTTTACCTTTTACTGAGCGGTTGCC 2974
QY 22 -----AlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrVal 38
Db 2975 GATACTGAATCACCGAGTAGAATACAGAACGAGTGTCTGGAGTTAGAACTATCCAGGTG 3034
QY 39 ValGlyLysSerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAsp 58
Db 3035 CAAGCCAAA-----CACGAGATCAGCAGCATGAC 3064
QY 59 LysIleIleSerGly-----AspThrLeuArgGlnLys 69
Db 3065 AATGAAGTCAACCGTTTGGGTAAAGTGTCAAAAGCAGTGAAGACATTGATAAAGAACTG 3124
QY 70 AlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGly 89
Db 3125 ATTTGAATATTCGCGATTTCACCCGTTATGATCCCGTATTTTCGGTGGAGCAGCGGA 3184
QY 90 GlyGlyAla---SerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValleu 108
Db 3185 CGTGGTCAACGTCAGGCTATGCAATCGTGGTGTTCACAGAACCCGCGTATGTTG 3244
QY 109 ---AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAsp-----His 124
Db 3245 GTGGACGCTTGGACAGCGCAGTCTCTATCTACCTTGAATCCGATGCCAACGGCGG 3304
QY 125 AlaIle---MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
Db 3305 GCGATTAAATGAATGATGAGATATTAATCAATTGAATTTGAGCAGCGGTCCAGT 3364
QY 144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIle 163
Db 3365 TCGGCAGAAATACGTTAGCGGTGCTTGGCGCGTGGTTCGAGGAGTTTCGACCAAGAACT 3424
QY 164 ProGlyLysMetProGluAsn-----GlyValSerGlyGluLeuGlyLeuArgLeu 180
Db 3425 GATGATGTGATTAAAGAGGGGCAAACTGGGGCTTGAACAGTAAACCGCTTACAGCAGC 3484
QY 181 SerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGly-----196
Db 3485 AAAAACACCCAGTATTAACCAATCCGTTGCCGCTGCTTCCGTTCCGCGCGGTTTTCAGAGT 3544
QY 197 -----LeuGlyLysAsnPheValLeuHis-----204
Db 3545 TTGGCGATTTTACCCTATCGTAAAGGTAAAGAAACCCCGCTGCATCTCTGCTGCCAAGAA 3604
QY 205 -----ThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaVal 217
Db 3605 ATACAACATACCTACCAACCATTTGGAAGGTTATTTTATTCGG-----TATGAGTT 3655
QY 218 ProArgTyrArgAsnLeuLys-----224
Db 3656 GACCAAAACCCGAAACGAAAGCCTGTTCTGGCAATGCGTATTTATATATCTTGTCCGATGAA 3715
QY 225 -----ArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaVal---241
Db 3716 TGCTCTAATCTAAGTATCCGAGTTCGTCATGCCAAGGCCCAAGCAAGATAGGTGGGT 3775
QY 242 -----LeuGlyTyr-----ArgLysArg 247
```

Db 3776 GCCCGGAGACAATCCTAATTGGACGCCGCCGAGACGAGCAGCGCTCTCTAAATGGCG 3835
 QY 248 PheTyrArgXThrTyrSer-----AspArgArgAspGlnTyr 260
 Db 3836 TATCCGACACGATACCGCCTCTGCCAAGATTATACGGGTCTCGACCGCATCAGCCCTAAT 3895
 QY 261 GlyLeuProAlaHisSerHisGlnTyr----- 269
 Db 3896 CCGATGGACTACCAAGTCACTCTTCTCTCTGGAAGGTGGTTACCGCTTGCGCCTAAC 3955
 QY 269 ----- 269
 Db 3956 CATTATGTCCGGCGGTGTGGAAACATACGAACGACGCGTTACGATATCCGTGATATGACG 4015
 QY 270 -----AspAspCysHisAlaAspIleIleTyrGlnLysSer 281
 Db 4016 CAACGGCGGTATTACAGAAAGAGGATATCTGCCACAGCGATCCAGTTGCCAAACGTG 4075
 QY 282 LeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAsp 301
 Db 4076 GATAAAAATGAGACGGACAAAGGTAATTCGGTATCACGTTGACTGATATCCTTTGGAC 4135
 QY 302 -----TyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp----- 314
 Db 4136 GGTTCGTATATATGCGCGCAATCAAGCTCGTGGCGTCCGGTACGACGGCGGTAAATTT 4195
 QY 315 ---AspAspAlaHisAlaHisAsnGly----- 324
 Db 4196 TTTAATGAACGCCATACGAAAATCGCTCGGTATCTTTTACCGCTATGAGATCCCGAT 4255
 QY 325 ---LysProTrp-----IleAspLeuArgAsnLysArgTyrGlu 336
 Db 4256 AAAAAATCTTGGCAGATAGCTTGACCTTGAGTATTGACCGCCCAAGATCTCAAACTGTCG 4315
 QY 337 LeuArgAlaGlnTrpLys-----GlnProPheProGlyPheGluAlaLeuArgValHis 354
 Db 4316 AGCGGTATCCATTGGACGTATGACCGGATATCTCATGTGGCACGCTTGGCGTCCGACG 4375
 QY 355 LeuAsn-----ArgAsnAspTyrHisAspGluLys 365
 Db 4376 TTGGACAAACCTTGGTCTAATTACCGTACCGAGAAAACGATTATCAAGACGACTCAAT 4435
 QY 366 AlaGlyAsp---AlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGlu 384
 Db 4436 CTGGACAAATTCATTTGGGAAAAAACTTTAATCTGGGCTTTTACCACGCGATTAAGTGAAT 4495
 QY 385 LeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGln 404
 Db 4496 ATC-----GCCGCGCGCTTTGGTACA 4516
 QY 405 LysSerSerAlaLeu-----SerAlaThrSerGluAlaValLysGlnProMetLeu 421
 Db 4517 CATCGCTCCACCTTACAACTACGCGACTTATATGCTGAATATGTCACTTGGCCACCGTAT 4576
 QY 422 LeuAspAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAspAsn 441
 Db 4577 ACAGAGGAAAAAGTG-----TATGGCGAAGATAATAAGTCAAAACAAAT 4621
 QY 442 PheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLys 461
 Db 4622 CCGACAGCAGAA-----GAAAAAGAGAA-----TTACATACGCGCAAT 4660
 QY 462 AlaLeuIleAspArgGluAsnTyrTyrLysGlnPro---LeuProAspLeu----- 477
 Db 4661 GGTTCCTTATGACAACTCGGTATATAGACGTAAAAACACGCGGMAATTAATAAACTGTC 4720
 QY 478 -----GlyAlaHisArgGln-----ThrAla 484
 Db 4721 AATGGGTGCAATGAGACAGCGCGATTAACCGTGACTGCTGCCACGTGTGATTACGGGC 4780
 QY 485 ArgSerPheAlaLeuSer----- 490
 Db 4781 AGACAGTATTACCTTGGCTTGGTAAACCATATTGCTTTGGTGAATGGCAGACTTGGGG 4840

QY 490 ----- 490
 Db 4841 TTGGCGGTCCGGTACGACAACCATACCTTCCGCTCGAATGACCCGTGGACCCAAAGTGCG 4900
 QY 491 -----GlyAsnTrpTyrPhe-----ThrProGlnHisLys 500
 Db 4901 AACTACCAACAATGGTGGTGAATCGGGCGGTGAGCCTCAAAACCAACCGCCACTTTGTC 4960
 QY 501 LeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyr----- 518
 Db 4961 GTGTCTTACCGTGTCTCCAGCGGTTTCCGCTGCCCGCTTTTATGAGCTGTACGGCGTG 5020
 QY 519 -----AlaHisGlyLys---HisValAlaThrAsnThrPheGluValGlyAsnLys 534
 Db 5021 CGTACGGGGCTTCTGTTAAAGACATCCACTCACAAAAAGAGTCTTTGACGCCGTAA 5080
 QY 535 HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArg 554
 Db 5081 CCGTTGAAAAGCGAAAAGCCCTTTAACCAAGAAATTTGGTTGGCCGTTCAGGGCGATT 5140
 QY 555 TrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThr 574
 Db 5141 GGTGTGATAGAGACAGTCTTCTTCCAAAAACAACATAAACAACCTGCTT----- 5188
 QY 575 LeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArg 594
 Db 5189 -----GCCCGTGCAGATAAATATGTCGAGGAGTTGGTTATGTATGTA----- 5227
 QY 595 TyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGlu----- 606
 Db 5228 -----ACCGATTTTACAAACCCACGATGTCAAACTCAACGGTATCAAT 5272
 QY 607 -----GlyGluIleTyrPheLys-----ProThrProArgTyr--- 617
 Db 5273 ATCTTGGGTAGAATCTACTGGGAAGCATCAGCGATAGGCTGCTGGAAGCTTGTATTCC 5332
 QY 618 -----ArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsn 632
 Db 5333 ACATTGCTTACAAACCGTATCAATATCAAGACGCAAAATTTGCACGCAATTTTACCAT 5392
 QY 633 Leu-----ProSerLeu-----ProGlyArgGluAspAlaTyrGlyAsnArg 646
 Db 5393 GTGTCTGAGCGACATTTGGAAGCGTCAACCGGACGCTATTATGCA----- 5440
 QY 647 ProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPhe 666
 Db 5441 ---AGTATCGCTATGATGAC-----CCTGAGGGCGAGATGGGCGCTT 5479
 QY 667 HisLeu-----LysAlaSerLeuThrAspArgIleAspAlaAsnLeuAsp 681
 Db 5480 AATTTAAGCGCACCTTACTCTCAAGCCAAACACGTCAGGAAGTGGTCGGCGAA----- 5533
 QY 682 TyrTyrArgValPheAlaGlnAsn-----LysLeuAlaArgTyrGluThrArgThr 698
 Db 5534 -----AAAGTTCGGCAGGTGGCAGCATTAACCGACGATCAACAGCAACGCACT 5587
 QY 699 ProGlyHisHisMetLeuLeuGlyAlaAsnTyrArg----- 711
 Db 5588 CGTCTTGGTATATTTATGATTTTACGCGCATCTACACTTGGAAAAGAAAAATTCACGTTG 5647
 QY 712 -----ArgAsnThrArgTyrGlyGluTyr---AsnTyrTyrVal 723
 Db 5648 AGACCGCGTATCTATAATTTAAACCAATCGTAATATAGCATGGAAAGTGTGGTGTAG 5707
 QY 724 LysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThr 743
 Db 5708 TCCGCTGCCAATGGGTCAATCAAGACCTAGGTACAGTTCGCGACGCTTTTGGC----- 5761
 QY 744 ProGlnMetGlyArgSerPheThrGlyValAsnValLysPhe 758
 Db 5762 ---GCACGGGCGCAGAACTTTTACCGTGAGTATGGAATGAAGTTT 5803

RESULT 33

US-10-238-075-1076
; Sequence 1076, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238, 075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1076
; LENGTH: 2046
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-1076

Alignment Scores:

Pred. No.:	1,59e-06	Length:	2046
Score:	155.50	Matches:	158
Percent Similarity:	31.8%	Conservative:	108
Best Local Similarity:	18.94%	Mismatches:	248
Query Match:	3.85%	Indels:	321
DB:	12	Gaps:	43

US-09-936-377-2 (1-758) x US-10-238-075-1076 (1-2046)

QY	51	LeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeuArgGlnLysAla	70
DB	130	CTCAGAAATGCCCGCCAGTGTCTCAGTACTTACCTCAGAACACTGCAGAAAAA	189
QY	71	ValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGly	90
DB	190	GTITCA-----GATCTGTCAGTGCAGTAAGAAGTGTGAAGGGATTAGTACTG	243
QY	91	GlyAlaSerAlaPro-----ValIleArgGlyGlnThrGlyArgIleLysValLeu	108
DB	244	GGGAATGAAAAACCGGATATCAGTATACGTGGTCTAAGTGGCGATTACACGCTG	303
QY	109	-----AsnHisGlyGlnThrGlyAspMet	117
DB	304	GTCTGATGACGCTCAGACGCTCGGGAATCCAGACCAACGACGCGCGGTTTGAA	363
QY	118	AlaAspPheSerProAsp---HisAlaIleMetValAspThrAlaLeuSerGlnVal	136
DB	364	GCGGATTTATCCCTCTCTGTGGAGCAAT-----GAACGCATT	402
QY	137	GluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal	156
DB	403	GAAGTATCGTGGCCCTATGTCTCCCTGTATGTTCTGATGCCATCGAGGGGTCATT	462
QY	157	AspValAlaAspGlyLysIleProGluLys-----	166
DB	463	AATATCAACCAAAACAGATTAAATACCAAAACATGGGATGGCGTACTTGGACTTGGGGG	522
QY	167	-----MetProGluAsnGly-----	171
DB	523	ATTATTCAGGAACATGGAAATTTGGTAATCAACCAAAATGACTTCTATCTGTGAGGC	582
QY	172	-----ValSerGlyGluLeuGlyLeuArgLeuSerSerGly---AsnLeuGluLysLeu	188
DB	583	CCATTGATTAAGGATAAATCTGTCTTCACTATATGGAGGAATGAATCGCAAGGAA	642
QY	189	ThrSerGlyGlyIleAsnIleGly-----	196
DB	643	GATAGT-----ATCTCTAGGGAAACACCGGAAAGATTAATGAATATAACGGCAACG	696
QY	197	-----LeuGlyLysAsnPheVal-----	202

DB	697	CTCCAGTTTACTCCGACTGAAGCCAGAAAGTTTGTGTTTGAATATGAAAAAATAACACG	756
QY	203	LeuHisThr-----GluGlyLeu-----TyrArgLysSerGlyAspTyrAla	216
DB	757	GTGCATACATTAAACACACTGGTGAGTCTCTCATGCTCGACTATCGGGA-----	807
QY	217	ValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGly	236
DB	808	-----AATCTTAAACAA-----CCAAACAGTAAAGAGAAACGATATTCA	849
QY	237	GlnHisArgAlaValLeuGlyTyr-----	244
DB	850	CGTAGTCACTGGTAGCAGCATGGAATGCCAGGCGGAAATACTGTCATCTCTGAAATTGCT	909
QY	245	-----ArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeu	262
DB	910	GTTTATCAGGAGAAAGTTATTCTGAGGTAAATCAGGTAAAGTAAAGATAATATAAT---	966
QY	263	ProAlaHisSerHisGlnTyrAspAspCysHisAlaAspIleIleTyrGln-----	279
DB	967	-----CATTGGGATCTTAATTACGAGTCAAGAAAA	996
QY	280	-----LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeu	295
DB	997	CGGAAATAACCAACACATCATAGTCAAAAGTACGCGCATTTCTGCCG-----	1047
QY	296	ThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAsp	315
DB	1048	-----GAAATGTA-----CTGACATCGGA-----	1068
QY	316	AspAlaHisAlaHisAsnGlyLysProTyrIleAspLeuArgAsnLysArgTyr	335
DB	1068	-----	1068
QY	336	GluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu	355
DB	1068	-----	1068
QY	356	AsnArgAsnAspTyrHisAspGluLysAlaGlyAspAlaValGluAsnPheAsn	375
DB	1069	-----GGTCAATTCAGCATCGAGACTCCGTCGATGACTCAGCCCGGTAATAAACG	1122
QY	376	AsnGlnThrGlnAlaAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly	395
DB	1123	ACGAACACAGCTCTGTTCAATTAACAGAAAGCTGTTTATAGAAAT-----	1173
QY	396	SerTyrGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla	415
DB	1174	-----GAATATGACGACACGATTTCTCGCCCTGACTGGA-----	1209
QY	416	ValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyValGlu	435
DB	1210	-----GGACTGCTCTCGATAAT-----CATGAAATCTATGGC-----	1242
QY	436	GlnAlaAsnTyr-----AspAsnPheThrLeu	444
DB	1243	-----AGTTACTGGATCCAGATTGTACGCTGTTTATAACCTGACCGCATATCTC	1299
QY	445	GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle	464
DB	1300	AAAGGGGGATCGCAAAAGCATTTCCGGCTCTCTCAATTCGT-----	1341
QY	465	AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla	484
DB	1342	-----GAGGTGAGTCTGGATTGGGATTTGGAACACTGACGCGAGGTGGT	1380
QY	485	ArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGln-----HisLys	500
DB	1381	GCCTCTATT---ATGTATGGAACAGGACCTGAAACCGGAGACAGTGTAAACGAGAG	1437
QY	501	LeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis	520
DB	1438	ATCGGTATTATTATAGTAATGATAGTGGTGTTCGCGAGCGGACGCTGTTTAATACT	1497

Qy	521	GlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys-----	534
Db	1498	GAITTTTAAAAAATAAGTTGACCACTTACGATATAGGTACAAAAGATCCAGTCACCGGGTTA	1557
Qy	535	-----HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu	548
Db	1558	AACACTTTTATTATGATATGATGAGTGGAGCAAAATATCAGAGGGGTGGAGCTTGCACCT	1617
Qy	549	GlyTyrGlu--GlyAspArgTtPginTyrAsnLeuAlaLeuTyrArgAsnArgPheGly	567
Db	1618	CAGATTCTCTGTATGATAAATGGCATGTATCT-----GCA	1653
Qy	568	AsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAsp	587
Db	1654	AACATAT-----ACATTTACTGACTCTCGT-----CGAAAAAGTGGATCGCAA	1695
Qy	588	SerGluMetLysLeuValArgTyrAsnGlnSerGlyValaAspPheTyrGlyAlaGluGly	607
Db	1696	AGT-----CTCAATGGCAAGTCGCTGAAGGGGAACCT---	1728
Qy	608	GluIleTyrPheLysProThrProArgTyrArgIleGly-----ValSerGlyAspTyr	625
Db	1729	-----CTGGAAAGAACTCCCCAGACATCGACCAATCAAAATCGGAATGGGATAC	1779
Qy	626	ValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsn	645
Db	1780	ACTCAGGAT--ATTACATTTTATTCATCTCTG-----AATTATACGGGAAAA	1824
Qy	646	ArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGly	665
Db	1825	CAAAATCTGGGCAGCACAA-----AGAAATGGTGCTAAGGTTCGCCGCTTCGTAATGGA	1878
Qy	666	PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal	685
Db	1879	TTC-----ACATCTATGATATATGCTCAAATACCCAG-----	1911
Qy	686	PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsn	705
Db	1912	ATTCTGCCAGACACGCTGATTAAITTTGCC-----GTTCTTAAC	1950
Qy	706	LeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTrpAsnTrpTyrValLysAl	725
Db	1951	GTCAC-AGACAGAAAGACGAGGATATCGATACCATTTGATGTTAACTGGCAGGCTC-----	2004
Qy	725	aAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerPheLeuSerAspThrProGly	745
Db	2005	-----GA	2006
Qy	745	nMetGlyArgSerPheThrGlyGlyValAsnValLysPhe	758
Db	2007	TGAAGGACCGCGTTATTTGGGCTTAATGTAAGTAGTATCTCTTC	2046

```

RESULT 34
US-10-240-218-5
; Sequence 5, Application US/10240218
; Publication No. US20030186848A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: Goringe, Andrew Richard
; APPLICANT: Hudson, Michael John
; APPLICANT: Matheson, Mary Anne
; APPLICANT: Robinson, Andrew
; APPLICANT: West, David McKay
; TITLE OF INVENTION: Recombinant iron uptake proteins
; FILE REFERENCE: GWS/22057
; CURRENT APPLICATION NUMBER: US/10/240,218
; CURRENT FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: GB 0007433.6
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5

```

:	LENGTH:	2748			
:	TYPE:	DNA			
:	ORGANISM:	Neisseria meningitidis			
<hr/>					
US-10-240-218-5					
<hr/>					
Alignment Scores:					
Pred. No.:	2_47e-06	Length:	2748		
Score:	155.50	Matches:	196		
Percent Similarity:	32.90%	Conservative:	136		
Best Local Similarity:	19.43%	Mismatches:	333		
Query Match:	3.85%	Indels:	346		
DB:	12	Gaps:	49		
<hr/>					
US-09-936-377-2 (1-758) x US-10-240-218-5 (1-2748)					
Qy	2	AlaGlnThrThrLeuLysProIleValLeuSerIleLeuLeuLeuLeuAsn-ThrProLeuLeu	21		
Db	3	GCAACAGCAACATTGGTTCGGATTCATATTTATATGCCTGTCTTTAATGACTCGCGTGCC	62		
Qy	21	uaalGlhAlahslGluThrGlu-----GlnSerValGlyLeuGluThrValTh	37		
Db	63	CGCTTATGCGAAATGTGAAGCCGCCAGACACAGAAAAACAGTTGNATACCATAACA	122		
Qy	37	rValVaIdcLyIsSer-----ArgProArgAlaThrSergLyLeuLeuHisTh	53		
Db	123	GGTAAGAAGCCAAAACAGAAAAACCCGCGGATAACAGAAATAACCGGGCTG----	174		
Qy	53	xSerThrAlaSerAspLysllelle---SerGlyAspThrLeu---ArgGlnLysAlaVa	71		
Db	175	-----GGCAAGTTGGTCCAAGTCTTCCGATACGCTAAGTAAGAACACAGGTTTT	221		
Qy	71	lAsnLeuClYAspAlaLeuAspGlyValProGlyVileHisAlaSerGlnTyrgLyGly	91		
Db	222	GAATTCACGACCTCGACCGCTTATGATCCGGTATTGCCGTGGTCGAACAGSGTCGGG	281		
Qy	91	yAlaSerAlaProVal---lleArgGlyGlnThrGlyArgIleLysVal-----	107		
Db	282	CGAAGTTCGGCTATTCAAATACGGCGCATGATAAAAAACCGCGTTCTCTTAACGGTGA	341		
Qy	108	-----LeuAsnHishi	111		
Db	342	CGGCGTTTCGAAATACAGTCTCTACACGCGCAGCGGCATTTGGCGGACGAGCACGC	401		
Qy	111	sGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAl	131		
Db	402	GGCGAGCAGCGCGCAATCAATGAATCAGATGAAACGTC-----	444		
Qy	131	aLeuSerGlnGlnValGluilleLeuArgGlyProValThrLeuLeuTyrserserGlyAs	151		
Db	445	-----AAAGCTGCGAAATCAGCAAAAGGCTCAAACTCGGTGCAACAGGACGCGGC	497		
Qy	151	nValAlaGlyLeuVal-----AspValalaaspGlyLysIleProGluLys--	166		
Db	498	ATTGCGGGCTCGGTGCGATTTCAAACCAAAACCCCGCACGATGTTATCGGGAAGGCGA	557		
Qy	167	-----MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSe	182		
Db	558	GCAGTGGGCATTCAGATGAACCCGCTATTCCGGCAA-----AACCG	602		
Qy	182	rglyAsnLeuGluLysLeuThrSerGlyglyIleAsnIleGlyLeuGlyLysAsnPheVa	202		
Db	603	GGGGCTTACCAATCCATCGCGCTGGCGGG---CGCATCGCGGTGCGGAGGCTTTGCT	659		
Qy	202	lleuHisThrGluGlyLeuTyArgLysSerclYasPTyr-----	215		
Db	660	GATCCACACCGG-----CGCGCGCGGGGGAAATCCGCGCCCCAGAAATGTCAGG	710		
Qy	216	----AlaValProArgTyArgAsnLeuLysArgLeuProAspSer-----	229		
Db	711	ACGGCGGTTACAGAGCTTTAACAGGCTGTGTCCGGTTGNAAGACAGACGAATTACGCTTA	770		
Qy	229	-----	229		

RESULT 35

```

US-10-238-075-1049
; Sequence 1049, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1049
; LENGTH: 12943
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-1049

Alignment Scores:
Pred. No.: 2,52e-05 Length: 12943
Score: 155.50 Matches: 158
Percent Similarity: 31.89% Conservative: 108
Best Local Similarity: 18.94% Mismatches: 248
Query Match: 3.85% Indels: 321
DB: 12 Gaps: 43

US-09-936-377-2 (1-758) x US-10-238-075-1049 (1-12943)

```

US-09-936-377-2 (1-758) x US-10-238-075-1049 (1-12943)

Qy	51	LeuHisThrSerThrAlaSerAspLysLleIleSerGlyAspThrIleuArgGlnLysala	70
Db	10842	CTCAGAAATCCCGCCGCACTGCTCAGTCATTACTTCAGAACAACTCAAAAAACCG	10901
Qy	71	ValAsnLeuGlyAspAlaLeuAspGlyValProGlyLleHisAlaSerGlnIleGlyGly	90
Db	10902	GTITCA-----GATCTGGTCGATGACGTAAAGATGTTGAAGGGATTAGTACTCGT	10955
Qy	91	GlyAlaSerAlaPro-----ValIleArgGlyGlnThrGlyArgArgIleLysValLeu	108
Db	10956	GGGAATGAAAAACCGGATATCAGTATACGTGCTGAAGTGGGGAATTACACGCTGATTCTG	11015
Qy	109	-----AsnHisHisGlyGlnThrGlyAspMet	117
Db	11016	GTCGATGACGACGTCTACAGCGTGGGAATCCAGACCAACGACGCGCGGTTTGA	11075
Qy	118	AlaAspPheSerProAsp---HisAlaIleMetValAspThrAlaLeuSerGlnGlnVal	136
Db	11076	GCCGGATTTATCCCTCTCTGTGGAGCAATT-----GAACGCATT	11114
Qy	137	GluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal	156
Db	11115	GAAGTGATCCGTGGCCCTATGTCTCCCTGTATGTTCTCATGCCATCGAGGGGTCAAT	11174
Qy	157	AspValAlaAspGlyLysLleProGluLys-----	166
Db	11175	AATATCATATAACAAACCAAGTTAATAACCAACATGGATGGCGTACTTGCACTTGGGGGG	11234
Qy	167	--MetProGluAsnGly-----	171
Db	11235	ATTATTGGAACATGGGAATTTGGTAACTCAACCAATATGACTTCTATCTGTACGGC	11294
Qy	172	-----ValSerGlyGluLeuGlyLeuArgLeuSerSerGly---AsnLeuGluLysLeu	188
Db	11295	CCATTGATTAGGATAAATCTGGTCTTTCAGCTATATGGAGGAATGAATCATGCAAGGAA	11354
Qy	189	ThrSerGlyGlyLleAsnIleGly-----	196
Db	11355	GATAGT-----ATCTCTCAGGGAACACCGGCAAAAGATATTAAGAATATAACGGCAACG	11408
Qy	197	-----LeuGlyLysAsnPheVal	202

Db	11409	CTCCAGTTTACTCCGACTGAAGCCAGCAAGATTGTTTTTGAATATGGAAAAATAACCCAG	11466
Qy	203	LeuHisThr-----GluGlyLeu-----TyrArgLysSerGlyAspTyrAla	216
Db	11469	GTGCATACATTAACACCTCGTGAGTCTCTCGATGCTCGACTATGCGGGGA	11519
Qy	217	ValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaasnGly	236
Db	11520	-----AATCTTAACAA-----CCAAACAGTAAAGAGAAACGCCATTAATCA	11561
Qy	237	GlnHisArgAlaValLeuGlyTyr-----	244
Db	11562	CGTAGTCACCTGGGTACGACGATGAATGCCAGGCGAAATACTGCATCTGAAATTGCT	11621
Qy	245	-----ArgLysArgPheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeu	262
Db	11622	GTTTATCAGGAGAAAGTTATTTCGTGAGGTAAATCAGGTAAAGATATAATATAAT	11678
Qy	263	ProAlaHisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGln-----	279
Db	11679	-----CATTGGATCTTAATTACGAGTCAAGAAA	11708
Qy	280	-----LysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeuLeu	295
Db	11709	CCGGAAATAACCAACAATCATAGATGCAAAAGTCACGCGCATTTCTGCGC	11759
Qy	296	ThrGluClnAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHisAspAspAsp	315
Db	11760	-----GAAATGTA-----CTGACCATCGGA	11780
Qy	316	AspAlaHisAlaHisAlaHisGlnLysProTyrIleAspLeuArgAsnLysArgTyr	335
Db	11780	-----	11780
Qy	336	GluLeuArgAlaGluTyrLysGlnProPheProGlyPheGluAlaLeuArgValHisLeu	355
Db	11780	-----	11780
Qy	356	AsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsn	375
Db	11781	-----GGTCAATTCAGCATGCAGAGCTCCGTGATGACTCAGCCACGGGTAAAAACG	11834
Qy	376	AsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly	395
Db	11835	ACAGAAACACAGTCTGTTTCAATATAACAGAAAGCTGTTTTATAGAAAT	11885
Qy	396	SerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAla	415
Db	11886	-----GAAATGACGACACGGATTCTCTCGCCTGACTCGA	11921
Qy	416	ValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGlu	435
Db	11922	-----GGACTGCTGCTCGATAAT-----CATGAAATCTATGCG	11955
Qy	436	GlnAlaAsnTrp-----AspAsnPheThrLeu	444
Db	11955	---AGTTACTGGAATCCAGATTGCTACGCTGTTTATAACCTGACCGCATTAATCTCACACTC	12011
Qy	445	GluGlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIle	464
Db	12012	AAAGGGGGATCGCAAAAGCATTTTCGGGCTCTTCAATTGCT	12055
Qy	465	AspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAla	484
Db	12054	-----GAGGTGAGTCCTCGAATTTGGAACTGACGACGCGAGGTGCT	12095
Qy	485	ArgSerPheAlaLeuSerGlyAsnTyrTyrPheThrProGln-----HisLys	500
Db	12093	GCCTCTATT---ATGTATGGAACACAGGCACTGAAACCGGAGACCAAGTAAACCGAAGAG	12149
Qy	501	LeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHis	520
Db	12150	ATCCGTATATTATAGTAATGATAGTGGTTTTTTCGGCGACGCGCGTGTATTATCT	12209

```
QY 521 GlyIysHisValAlaThrAsnThrPheGluValGlyAsnLys----- 534
Db 12210 GATTTTAAATAAGTACCAAGTTACGATATAGTACAAAAGATCCAGTCACCGGTTA 12269
QY 535 -----HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu 548
Db 12270 AACATTTTATTATGATAATAGTGGAGGCAATATCAGAGGGGTGGAGCTTCAACT 12329
QY 549 GlyTyrGlu--GlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGly 567
Db 12330 CAGATTCCTGTATGATAAATGGCATGTATCT-----GCA 12365
QY 568 AsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAsp 587
Db 12366 AACTAT-----ACATTACTGACTCTCGT-----CGAAAAAGTATGACGAA 12407
QY 588 SerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGly 607
Db 12408 AGT-----CTCAATGGCAAGTCGCTGAAAGGGGAACCT--- 12440
QY 608 GluIleTyrPheLysProThrProArgTyrArgIleGly-----ValSerGlyAspTyr 625
Db 12441 -----CTGAAAGAACTCCAGACATCGAGCCAAATGCAAAACTGGAATGGGATTAC 12491
QY 626 ValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsn 645
Db 12492 ACTCAGAT---ATTACATTTTATTCTCATCTCTG-----AATTATACGGGAAAA 12536
QY 646 ArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGly 665
Db 12537 CAAATCTGGCGACGACAA-----AGAAATGGTGCTAAGTTCCTCCCGCTGTAATGGA 12590
QY 666 PheHisLeuLysAlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgVal 685
Db 12591 TTC-----ACATCTCTGATATTGGTCTAAATTTACCAG----- 12623
QY 686 PheAlaGlnAsnLysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsn 705
Db 12624 ATTCTGCCAGACGCTGATTAATTTTGGC-----GTCTTAAC 12662
QY 706 LeuGlyAlaAsnTyrArgArgAsnThrArgTyrGlyGluTyrAsnTyrValLysAl 725
Db 12663 GTCAC-AGACAGAAAGACGAGGATATCGATACATGTATGTTAATCGCAGTCTC----- 12716
QY 725 aAspAsnLeuLeuAsnGlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGl 745
Db 12717 -----GA 12718
QY 745 nMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 12719 TGAAGGACGCGCTTATGGGCTAATGTAAAGATATCCTTC 12758
```

RESULT 36

US-10-043-344-1

Sequence 1, Application US/10043344
Publication No. US20030088086A1
GENERAL INFORMATION:
APPLICANT: Loomis, Sheena M.
APPLICANT: Harkness, Robin E.
APPLICANT: Schryvers, Anthony B.
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Murdin, Andrew D.
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
FILE REFERENCE: 1038-1221 MIS
CURRENT APPLICATION NUMBER: US/10/043,344
CURRENT FILING DATE: 2002-07-01
PRIOR APPLICATION NUMBER: 08/649,518
NUMBER OF SEQ ID NOS: 160
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1
LENGTH: 4699
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: CDS
LOCATION: (10)..(1940)
FEATURE:
NAME/KEY: CDS
LOCATION: (1957)..(4696)
US-10-043-344-1

Alignment Scores:
Pred. No.: 6,278-06 Length: 4699
Score: 155.00 Matches: 188
Percent Similarity: 32.33% Conservative: 144
Best Local Similarity: 17.41% Mismatches: 309
Query Match: 3.84% Indels: 344
DB: 14 Gaps: 48

US-09-936-377-2 (1-758) x US-10-043-344-1 (1-4699)

```
QY 5 ThrLeuLysProIle-----ValLeuSerIleLeuLeuIleAsnThrProLeu 20
Db 1961 ACTAAAAAACCTTATTTTCGCTAAGTATTATTTCTGTCTTTTAATTCATGCTATGTA 2020
QY 21 LeuAlaGlnAlaHis-----GluThrGluGlnSerValGly----- 32
Db 2021 AAAAGCAGAAACTCAAAGTATAAAAGATACAAAGAAAGCTATATCATCTGAAGTGGACACT 2080
QY 33 -----LeuGluThrValThrValValGly---LysSerArgPro 44
Db 2081 CAAAGTACAGAGATTTCAGAACTAGAACTAATCTCAGTCACTCGACGAAAAGTTAGAGAT 2140
QY 45 ArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle-----Ser 62
Db 2141 CGTAAAGATAAATGAAGTA-----ACTGGACTTGGCAAAATATATAAAACTAGT 2188
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 2189 GAAAGTATCAGCGAGACAAAGTATTAAATATATTCGTGATCTAACACGCTATGATCCAGG 2248
QY 83 IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal---IleArgGlyGlnThr 101
Db 2249 ATTTCAGTTGTAGACAGAGTCGCGGTGCAAGTTCGGATATTCTATTCTGTGTATGGAC 2308
QY 102 GlyArgArgIleLysValLeu----- 108
Db 2309 AGAAATAGAGTTCTCTTTTATAGTAGTGGTTTACCTCAAACGCAATCTTATGTAGTGCAG 2368
QY 109 -----AsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro 122
Db 2369 AGCCCTTTTATGCTCGTTCAGGATATTTCTGGCAGTGGTGCATTAATTAATTAATTAAT 2428
QY 123 AspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyPro 142
Db 2429 GAAATGTA-----AAGCCCTCGAAATAAGCAAGGGGGG 2464
QY 143 ValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuVal----- 156
Db 2465 AGTTCTTCTGAGTATGTAATGAGCACTAGTGGTCTCTGTAAACATTTTCAAGCAAAATCA 2524
QY 157 -----AspValAlaAspGly-----LysIleProGluLys 166
Db 2525 GCAGCCGATATCTTAGAAGGAGACAAATCATCGGGAATTCAAACTAAATATGCTTATTC 2584
QY 167 MetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGlu 186
Db 2585 AGCAAAATAAAGCCTTTTACCCTATTTCTTCTAGCTGTACAGAGAAACAAAGTGGATTGAA 2644
QY 187 LysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsn----- 200
Db 2645 -----GGGGTGGCCATTTTACACTCACCGAATTCATTAATTGAACCAAGTC 2699
```

QY 201 -----PheValLeuHisThrGlu--- 206
 Db 2690 CATAAGATGCATTAAGCGGTGCAAGTATGATCGATTATCGCCCAACAGAGGAT 2749
 QY 207 ---GlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArg 225
 Db 2750 CAATCGCATCTTGTGATGCAAGATGAGTGTCTAGATGGTTATGACAAGTGTAAGAACT 2809
 QY 226 LeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeu--- 242
 Db 2810 TCACCCAAACGACCTCGACTTATCCACCCAAAGAAACCGTAAGCGTTTCAGATTAT 2869
 QY 243 -----GlyTyrPArgLys 246
 Db 2870 ACGGGGCTAACCGTATCAAACTTAATCAATGAATATGAAGCCAGCTCTGG--- 2923
 QY 247 ArgPheTyrArgArgThrTyrSerAspArgAspGlnTyr-----GlyLeuProAla 264
 Db 2924 ---TTTAAAGAGAGGATTATCATTTTCTGAAACACATATATGTGTGGTAT--- 2974
 QY 265 HisSerHisGluTyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsn 284
 Db 2975 -----TTTGAATTCACACACAAATAATTTGATATC----- 3004
 QY 285 LysArgTyrLeuGlnLeuTyrProHisLeuLeuThrGluLeuAspValAspTyrAspAsn 304
 Db 3005 ---CGTGATATGACATTTCCCGCTTATTTAAGGCCAACAGAGGATTTACAA--- 3058
 QY 305 ProGlyLeuSerCysGlyPheHisAspAspAlaHisAlaHisAlaHisAsnGly 324
 Db 3059 -----AGTCGCCCTTTTCCAAAGCAAGATTTATGTGTCATATCAACATTTGGT 3109
 QY 325 LysProTyrIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnPro 344
 Db 3110 -----GATGGCAGAGCGGCTTAAATAT----- 3130
 QY 345 PheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGlu 364
 Db 3131 -----GCAAGTGGGCTTTATTTTCGATGAACACCATAGAAAAACAG 3169
 QY 365 LysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGlu 384
 Db 3170 CGTGTAGT-----ATTGAATATATTTACGAAATAAAGAACAAAGCGGCATCATTTGAC 3223
 QY 385 LeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyrLeuGlyGln 404
 Db 3223 ----- 3223
 QY 405 LysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsn 424
 Db 3224 ---AAAGCGGTGTTAAGTCTAAATCAACAAACATCATCTTGACAGTTATATGCGACATA 3280
 QY 425 LysVal-----GlnHis 428
 Db 3281 CGCATTTGAGCTTTTATCCATATCCAACTAAGATTCGCGCCCAACACTTGATTAACCT 3340
 QY 429 TyrSerPhePheGlyValGlu----- 435
 Db 3341 TATTCATCATCTCTGATAGAAATGTTTATAAAGAAAAACATAACATGTTGCAATTG 3400
 QY 436 -----GlnAlaAsnTyr----- 439
 Db 3401 AATTAGAGAAAAAATTCACAAATTTGGCTTACTCATCAAAATGGCTTCAATCTTGGT 3460
 QY 440 ---AspAsnPheThr-----LeuGlu 445
 Db 3461 TTTGATGACTTTACTTCCGCACTTCAGCATAAAGATTATTTAACTCGACGTGTATCGCT 3520
 QY 446 GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAsp 465
 Db 3521 ACGGCAAGTAGTATTTTCAGAGAAACCGTGGTGAAGCAAGAAAGAAATGGTTTA----- 3571

RESULT 37

QY 466 ArgGluAsnTyrTyrLysGlnProLeuPro-----AspLeuGly 478
 Db 3572 CAATCAAGTCCCTTACTTATACCCAAACAGAGAGTGTGTAGAGAGATCTTTGT 3631
 QY 479 AlaHisArgGlnThrAlaArgSerPheAla-----LeuSerGly--- 491
 Db 3632 AATTATCAAGTAAAGTCCCTTAATACAGTGACTGTAAGTGGGTAAATTAAGGAAAA 3691
 QY 492 AsnTyrTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeu 511
 Db 3692 AATTATTTATTCAGCAGCAATAATATGCGCATTA----- 3727
 QY 512 ProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluVal 531
 Db 3728 -----GGGAATACGTT-----GATTTAGGTTTA 3751
 QY 532 GlyAsnLysHis---LeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGly--- 549
 Db 3752 GGTATGAGGTATGACGTATCTCTGACAAAGCTAATGAATCAACTATTAGTTGGTAAA 3811
 QY 550 TyrGluGlyAspArgTyrGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyr 569
 Db 3812 TTTAAAAATTTCTCTTGGAACTACTGTTATTTGTAATAAACCAACGGAATGCTTGATCTT 3871
 QY 570 IleTyrAlaGlnThrLeuAsnAspGly---ArgGlyProLysSerIleGluAspAspSer 588
 Db 3872 TCTTAT-----CGCTTTTCTCTGATTTAGAAATCTTAGTTT-----GCT 3913
 QY 589 GluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGlu 608
 Db 3914 GAAATGATGTTGGCGGTATGGTGGCAAGATACCGATGTTTATATAGGTAAA----- 3967
 QY 609 IleTyrPheLysPro-----ThrProArgTyrArgIleGlyValSerGlyAsp 624
 Db 3968 -----TTTAAGCTCGAAACATCTCGTAACCAAGATTTGGTCTCGCTCTAAAGGGAT 4021
 QY 625 TyrValArgGlyArgLeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGly 644
 Db 4022 TTTGGTATATGATGATCAGTCATTT-----AGTAATGTTATCGA 4063
 QY 645 AsnArgProPheIleAlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeu 664
 Db 4064 AATCTTATCGCTTTGCTGGAAGAACTTAGTAAAAATGGAACCTACTGGAAGGCAATTTAT 4123
 QY 665 GlyPheHis-----LeuLysAlaSerLeuThrAspArgIleAsp 677
 Db 4124 GGATATCATATGCAAAATGCAAAATGTAATGTTGGCGTAATATATACTGCGCAATTAGAT 4183
 QY 678 AlaAsn-----LeuAspTyrTyrArgValPheAlaGlnAsnLys 690
 Db 4184 TTTTATGTTTATGGAACGTTATCCCTACGTTGTTGATGCAACATTTGCTTATAACCGA 4243
 QY 691 LeuAla-----ArgTyr 694
 Db 4244 GTAAAGTTAAAGATCAAAAAATCAATGCTGTTTGTAGCTTCGTAAGCAGTTTATTTAT 4303
 QY 695 GluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThr 714
 Db 4304 GATGCCATTCAGCCAGCGCTTATATCATTTAGGCTATGATCATCCCAAGTAATACT 4363
 QY 715 ArgTyrGlyGluTyrAsnTyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyr 734
 Db 4364 -----TGGGAATTAAG-----ACAATGTTT 4384
 QY 735 AlaHisSerSerPheLeuSerAspThrProGlnMetGlyArgSerPheThrGly----- 752
 Db 4385 ACTCAATCAAAAGCAAAATCTCAAAATGAATTTGCTAGGAAAAACGTGTCATTTGGGTAAACAT 4444
 QY 753 GlyValAsnValLys 757
 Db 4445 TCAAGGAATGTAATA 4459

2609 AATGTAAGCCGCTCGAA---ATAAGCAAGGGGGGAGTCTTCTCTGAGTATGGTAATGGA 2666
180 LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGly--- 198
2666 GCAC TAGCTGGTCTCTGTAACATTTCAAGAACAAATCCGACGCCGATATCTTTAGAAGGAGAC 2735
199 LysAsnPheValLeuHisThrGluGlyLeuTyArg----- 210
2726 AAATCATGGGAATTTCAAATCTAAATATGCTTTATCAAGCAAAAATAAAGGCTTTACCCAT 2785
211 -----LysSerGlyAspTyr-----AlaValProArgTyArg 221
2786 TCCTTAGCTGTAGCAGCAAAACAAGGTGGATTTCAAGGGGTGCGCAITTTACACTCAACGA 2845
222 AsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaVal 241
2846 AATTCGGAGGAA-----ACCAAGTCCATAAAGATGCA 2878
242 LeuGlyTrpArgLysArgPheTyArg-----ArgThrTySerAspArgArgAspGln 259
2879 TTAAGAGCGGTACAAAGTATGAGGATTCATCCGCCACAACAGATAAATCTTCAGGATAC 2938
260 TyrGlyLeuProAlaHisSerHisGluTyArgAsp---CysHisAlaAspIleLeuTrp 278
2939 TTTGTGATACAAAGGTGAGTGTCCAAATGGTGATGACCAAGTGTGAGCC----- 2988
279 GlnLysSerLeuIleAsnLysArgTyLeuGlnLeuTyProHisLeuLeuThrGluGlu 298
2987 -----AAACCCCTGCAAAAGTTATCCCCCAAGCAACCGTAAGC 3028
299 AspValAspTyr-----AspAsnPro----- 305
3029 GTTTCAGATTATACGGGGCTTAACCGTATCAAACTCAATCCAAATGAATATATGAAGCCAG 3088
306 ---GlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsn 323
3089 TCTTGGTGTATTAAGAGGAGGTATCATTTTCTGAACAA-----CACTATATTCGT 3133
324 GlyLysProTrpIleAspLeuArgAsnLysArgTyGluLeuArgAlaGluTrpLysGln 343
3140 GGT-----AATTTTGAATTCACACACACAAAAATTGATATCCGT-----CATATG 3183
344 ProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyHisHisAsp 363
3185 ACATTTCCCGCTTAT-----TTAAGATCAACA 3211
364 GluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAla----- 381
3212 GAAAAACGGGATGATAGAACTGGCCCTTTTATCCAAACCAAGATTATGGTGCATATCAA 3271
382 ArgIleGluLeuArgHisGlnProIleGlyArg----- 392
3272 CGTATTGAG-----GATGCCGAGGCGTTTAACATATCGAAGTGGGCTTTAT 3311
393 -----LeuLysGlySerTrpGlyValGlnTyLeuGlyGlnLysSerSer 407
3317 TTCGATGAACACATAGAAAAACGGGTGTAGGTATTGAATATATTTACGAAATAAGAAC 3371
408 AlaLeuSerAlaThrSerGluAlaVal-----LysGlnProMetLeuLeuAsp 423
3377 AAAGCGGGCATTCATTGACAAACAGCAGTGTTAAGTGCTTAATCAACAAAAACATCATACTTGAC 3433
424 AsnLysValGlnHis----- 428
3437 AGTTATATGCGACATACCGATTGCACTCTTATCCTAATCCAAGTAAGAAATTCGCGCCG 3496
429 -----TySerPheGlyValGlu----- 435
3497 ACACCTGATAACCTTATTCATCTACTATCGTCTGTGATAGAAATGTTTATAAAGAAAAACAT 3551
436 -----GlnAlaAsnTrp----- 439
3557 AATATGTTCAATTTGAATTTAGAGAAAAAAATTCACAAAAATGGCTTACTCATCAAAAT 3611

QY 440 ----- AspAsnPhe ----- 442
Db 3617 GTCTCAATCTGGTTTGTGACCTTTACTTCAGCGCTTCAGCATAAAGATTATTAACT 3676
QY 443 ----- ThrLeuGluGlyValArgVal 450
Db 3677 CGACGTGTTACCGCTACGCCAAATATTATTTTCAGGCACAGTTGCTGTGTAAGCAAGAAAT 3736
QY 451 GluLysGlnLysAlaSerIleArgTyrAspLys ----- 461
Db 3737 GGTTCAGAAAACAAACCTTACTTATCTACCAAAACCAAAAGTAGATTTTGTAGGACACAGAT 3796
QY 462 ----- AlaLeuIleAsp 465
Db 3797 CATTCGTAATTATAAGGTAGTCCTCTAATTACAGCGACTGTAAAGTGGGTTAATTAAA 3856
QY 466 ArgGluAsnTyrTyr ----- LysGlnProLeuProAspLeu 477
Db 3857 GGGAAAAATTATTATTTTCGACGACGCAATATAATATGGCAATTAGGGAATACATTTGTTA 3916
QY 478 Gly ----- AlaHisArgGlnThrAla ----- 484
Db 3917 GGTTCAGTATTCGGTATGACGTATCTCGTACAAAAGCTAATGAATCACTATTAGTGTT 3976
QY 485 ----- ArgSerPheAlaLeuSerGlyAsnTyrPheThrProGlnHisLysLeu 501
Db 3977 GGTAAATTTTAAATTTCTCTTGGAACTACTGGTATTTGTCATAAAACCAACGGAATGGCTT 4036
QY 502 SerLeuThr ----- AlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAla 519
Db 4037 GATCTTTCTATCGCTTCTCTACTGGATTAGAAATCCCTAGTTTTCCTGCAATGTAT --- 4093
QY 520 HisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGlu 539
Db 4094 ---GGTTCGGTATCGTGGCAATAATAGCGATGTTTATGTAGGTAAATTTAAGCCTGAA 4150
QY 540 ArgSerAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeu 559
Db 4151 ACATCTCGTAAACAGAGTTTGTCTCGCTCTCTAAAGGGGATTTTGTATATTTAGATC 4210
QY 560 AlaLeuTyrArgAsnArgPheGlyAsnTyrIle ---TyrAlaGlnThrLeuAsn --- 576
Db 4211 AGTCATTTTAGTAATGCTTATCGAAATCTTATCGCCCTTTGCTGAAGACTTAGTAAAT 4270
QY 577 ---AspGlyArgGlyProLysSerIleGluAspSerGluMetLysLeuValArg 594
Db 4271 GGNACTACTGGAAGGGCAATTATGATATCATATATGCACAAAATGCAAAATTTAGTTGC 4330
QY 595 TyrAsn ---GlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPhe --- 611
Db 4331 GTAAATATACTCGCAATTAGATTTTAAATGTTTATGGTTTATGGAACGTTATCCCTACGTTGG 4390
QY 611 --- 611
Db 4391 TATGCAACATTTGCTTATACCGAGTAAAGTTAAAGATCAAAATCAATGCTGTTTG 4450
QY 612 --- LysProThrProArgTyrArgIleGly 620
Db 4451 GCCTCCGTAAGCAGTTATTATTATTGATGCAATTCAGCCAGC ---CGTTATATCAITGTT 4507
QY 621 ValSerGlyAsp ---TyrValArgGlyArg 629
Db 4508 TTAGGCTATGATCATCCAGTAAATACCTGCGGAATTAATACATGTTTACTCAATCAAAA 4567
QY 630 LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
Db 4568 GCAAAATCTCAAATGATTTGCTAGGACACGT ---GCATTGGGTAC --- 4612
QY 650 AlaGlnAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLys 669
Db 4613 ---AATTCAGGAATGTAAATCAACAAGAAACTTACTCGGGCATGGCATATC --- 4663

QY 670 AlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsn 689
Db 4664 ---TTAGATGTATCGGGTTATTACATGCGGAAT 4693
QY 690 LysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsn 709
Db 4694 AAA ---AATATTATGCTTCGATTAGGATATAT 4723
QY 710 TyrArgArgAsnThrArgTyrGlyLeuTrpAsnTrpTyrValLys ---AlaAspAsnLeu 728
Db 4724 AATTATTCAACTATCGCTATGTTACTTGGGAAGCGGTGCGTCAAAACAGACACAGGTGCG 4783
QY 729 LeuAsn ---GlnSerValTyrAlaHisSerSerPheLeuSerAspThrProGlnMet 746
Db 4784 GTCAATCAACATCAAAATGTTGGTAGCTATATCTCCTACGCGACATCA --- 4831
QY 747 GlyArgSerPheThrGlyValAsnValLysPhe 758
Db 4832 GGACGAAACTATACCTTAACATTAGAAATTC 4867
RESULT 38
US-09-332-226-1
; Sequence 1, Application US/09332226
; Patent No. US20020025318A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332,226
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/363,124
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3286 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:


```

Db      2185  ---AGTGGCGGATGTCGGCGGCTTCGCTACGAC----- 2220
Qy      468  AsnTyrTyrLysGlnProLeuProAspLeuGlyAla---HisArgGlnThrAlaArgSer 486
Db      2221  -----TACCGCAGCAGCATTCGGACAGCGCGCTTCACCGCAGCGCAGCGCACC 2274
Qy      487  PheAlaLeuSerGlyAsnTyrTyrPheThrProGlnHisLysLeuSerLeuThr----- 504
Db      2275  CTGTCCTCGAAGCGCGCATGCTCCTCAAACTTCGCGAGCTGGCTGGATTTGACTTACCGC 2334
Qy      505  AlaSerHisGlnGluArgLeuProSerThrGlnGlnLeuTyrAlaHisGlyLysHisVal 524
Db      2335  ACTTCAACCGGCTTCGCGCTCCCTCGTTTCGCGAATGTACGCG----- 2379
Qy      525  AlaThrAsnThrPheGluValGlyAsnLys-----HisLeuAsnLysGluArg 540
Db      2380  -----TGGCGGTGGCGGATATAATAAAAGCGCGTCAAAATCGATCCGGAATAA 2427
Qy      541  SerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAla 560
Db      2428  TCGTTCACAAAGACCGCGCATGCTGTGTTAAAGCGGATTTTCGCAACTTCGGAGCGCAAGT 2487
Qy      561  LeuTyrArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsnAsp 577
Db      2488  TGGTTCACATGCTACCGCGATGTGATTGTCGCGGTATGAAGCGCAATTAAGAC 2547
Qy      578  GlyArg-----GlyProLysSerIleGluAspSerGluMet----- 590
Db      2548  GGCAAGAACAAAGTCAAAAGCAACCGCGCTTACCTCAATGCCCAAGCGCGGATTACC 2607
Qy      591  -----LysLeu----- 592
Db      2608  GGCATCAATATTTGGGCAAAATCGATTGGAACGGCGTATGGGATAAATGCCCGAAGGT 2667
Qy      593  -----ValArgTyrAsnGlnSerGlyAlaAsp 601
Db      2668  TGGTATTCACATTTGCTGCTATATCGTGTGTCGCGCATCAAAAGCGCGCAGAC 2727
Qy      602  PheTyrGlyAlaGluGlyGluIleTyr-----PheLysProThr----- 614
Db      2728  CGCACCAGATATTCATCACACCTGTTGATGCCATCCAAACCTCGCGCTATGTCGTCGGC 2787
Qy      615  -----ProArgTyrArgIleGlyValSerGlyAspTyrTyrValArgGlyArg 629
Db      2788  TCGGCTATGACCAACCGAAGCAATGCGCGTGAACGGTATGCTGACTTATTCCTCAA 2847
Qy      630  LeuLysAsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIle 649
Db      2848  GCCAAGGAAATCACAGAGTTG-----TTGGCGAGCGCGCTTTGCTC 2889
Qy      650  AlaGlnAspAspGlnAsnAlaProArgValProAlaAlaArgLeuGlyPheHisLeuLys 669
Db      2890  AACGGCAACAGCGCAATACAAAGCCACCGCGCGTACCCGCTTGTGTATATT--- 2946
Qy      670  AlaSerLeuThrAspArgIleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsn 689
Db      2947  -----GTGACGTGTCC---GGTATTACCGGTT----- 2973
Qy      690  LysLeuAlaArgTyrGluThrArgThrProGlyHisHisMetLeuAsnLeuGlyAlaAsn 709
Db      2974  -----AAAAAACACTTCACCTCCGTCGCGCGGTGATC 3006
Qy      710  TyrArgArgAsnThrArgTyrGlyGluTrp---AsnTyrTyrValLysAlaAspAsnLeu 728
Db      3007  AACCTCTCAACACCGCTATGTTACTTGGGAAATGTGGCGCAAACTGCCGCGCGCGCA 3066
Qy      729  LeuAsnGlnSerValTyrAlaHis-SerSerPheLeuSerAspThrProGlnMetGlyArg 748
Db      3067  GTCAACCAA-----CACAAAATCTCGCGTATTACACCGGATATGCCGCGCCCC--GGCGG 3119
Qy      748  gSerPheThrGlyGlyValAsnValLysPhe 758

```

```

Db      3120  CAACTACACATTTAGCTTGGAAATGAAGTTC 3150
RESULT 39
US-09-332-226-3
Sequence 3, Application US/09332226
Patent No. US20020025318A1
GENERAL INFORMATION:
APPLICANT: Sparling, P. Frederick
APPLICANT: Cornelissen, Cynthia N.
TITLE OF INVENTION: Transferrin-Binding Proteins From
TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: InClone Systems Incorporated
STREET: 180 Varick Street
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/332,226
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/363,124
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/973,336
FILING DATE: 05-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/572,187
FILING DATE: 23-AUG-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gallagher, Thomas C.
REGISTRATION NUMBER: 37,066
REFERENCE/DOCKET NUMBER: SPA-1-PDC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 645-1405
TELEFAX: (212) 645-2054
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3537 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEetical: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Neisseria meningitidis
STRAIN: FAM18, FAM20, B16B6, group X and group W135
FEATURE:
NAME/KEY: CDS
LOCATION: 721..3450
NAME/KEY: mat-peptide
LOCATION: 793..3447
US-09-332-226-3
Alignment Scores:
Pred. No.: 2,4e-05
Score: 148.00
Percent Similarity: 32.20%
Best Local Similarity: 20.46%
Query Match: 3.67%
DB: 9
Length: 3537
Matches: 204
Conservative: 117
Mismatches: 324
Indels: 355
Gaps: 48

```



```

Qy 540 rSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuA 560
Db 2720 AATCCTTAATAGAGAGGAGGAGTATGATTTAAAGGGGACTTCGGCAATTTGGAAGCA 2779
Qy 560 laLeuTyrArgAsnArgPheGlyAsnTyrIle-----TyrAlaGlnThrLeuAsnA 577
Db 2780 GCTATTTCAACAATGCCATCGGACCTGATTCGATTCGATTCGATTCGATTCGATTCG 2839
Qy 577 spGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnG 597
Db 2840 ACGGCA----- 2847
Qy 597 lnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrProArgT 617
Db 2848 -----ACTTCGGCTTCGGGAC-----CCGGAT 2872
Qy 617 yArgIleGly-----ValSerGlyAspTyrValArgGlyArgLeuLys- 631
Db 2873 ACCGAATGGCCCAAAATGCACGGTAGTAGCCGGTATCAATATTTGGTAAATCGATT 2932
Qy 632 -----AsnLeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnA 646
Db 2933 GGCACGGCGTATGGGCGGTTGCCGACGGTTGATTCACCGCTTGCTAT---AACC 2989
Qy 646 rgProPheIleAlaGlnAspAspGlnAsnAlaProArg----- 658
Db 2990 GTATCAAGGTCAAGATGCCGATATACGCCGCCAGGAGCTTTGTAATTCATATCTCT 3049
Qy 659 -----ValProAlaAlaArg-----LeuGlyPhe----- 666
Db 3050 TTGATCCCGTCCACCTTCACGATATGATTGGTTTGGTTACGACCATCTCGACGGA 3109
Qy 667 -----HisLeuLysAlaSerLeuThrAspArg----- 675
Db 3110 TATGGGCATCAATACGATGTTTACTTATTCAGGCAAAATCTGTTGCAAGACTGCTCG 3169
Qy 676 -----lleAspAlaAsnLeuAspTyrTyrArgValPheAlaGlnAsnLysL 691
Db 3170 GCAGCCAGCGCTGTGTGAAGGATGATGCAATGCTAAAGAAAGAGCAGTCAACGCGGAC 3229
Qy 691 euAlaArgTyrGluThrArgThrProGly-----HisHisMetLeuA 705
Db 3230 GGCCTGGTATGTTACGGATGTTCCGGATATTACAATATCAAGAAACACCTGACCCCTG 3289
Qy 705 snLeuGlyAlaAsnTyrArgAsnThrArgTyrGlyGluTrp---AsnTrpTyrValL 724
Db 3290 GGCAGGTGTGTACAACTCTCACTACCGCTATGTTACTTGGGAAAATGTGCGGCAAA 3349
Qy 724 ysAlaAspAsnLeuLeuAsnGlnSerValTyrAlaHis-SerSerPheLeuSerAspThr 743
Db 3350 CTGCGCGCGCGCAGTCAACCA-----CACAAAATGTGCGGTTTACACCGGATATG 3403
Qy 744 ProGlnMetGlyArgSerPheThrGlyGlyValAsnValLysPhe 758
Db 3404 CCGCCCCC-GGCGGAAATACACATTTAGCTTGAATGAAGTTT 3447

```

RESULT 40

```

US-09-815-242-6035
; Sequence 6035, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

```

```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6035
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(2190)
US-09-815-242-6035

```

```

Alignment Scores:
Pred. No.: 6,03e-05 Length: 2190
Score: 141.50 Matches: 151
Percent Similarity: 32.36% Conservative: 93
Best Local Similarity: 20.03% Mismatches: 275
Query Match: 3.51% Indels: 235
DB: Gaps: 35

```

US-09-936-377-2 (1-758) x US-09-815-242-6035 (1-2190)

```

Qy 3 GlnThrThrLeuLysProIleValLeuSer-----lleLeuLeuIleAsnThrProLeu 20
Db 37 CAAGCCATCACCAACCGTCACCTACTTCCGGTTCGATAGCAGCTATTTACCTTCT 96
Qy 21 LeuAla---GlnAlaHisGluThrGluGlnSerVal-----GlyLeuGluThrVal 36
Db 97 GCGCTTTTGTGTCACCGACCTAGAAACCGTGATTTGTTGAGGTTTCAGCCAGCT 156
Qy 37 ThrValValGlyLysSerArgProArgAlaThrSer-----GlyLeuLeuHisThr 53
Db 157 CCAGATGATGCGGAAATGATACAGCGTAACGCTCTACCTCTCGCGGTACCAAAATGCAG 216
Qy 54 SerThrAlaSerAsp-----LysIleIleSerGlyAspThrLeuArg--- 67
Db 217 ATGACTCAACGTGATATTCCTCAGTCGGTCACTATTGTTAGCCAGCAGCGGATGGAAGAT 276
Qy 68 GlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGln 87
Db 277 CAGCAGTTACAAACGCTGGCGAAGTATGAGAAACACACGCTGGGATCAGCAAAAGTCAG 336
Qy 88 TyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysVal 107
Db 337 GCGGATTCGATCGTCTCTTTTATTATTCGCCGGA-----TTCCAGATC 381
Qy 108 LeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHis----- 124
Db 382 GATTAATAT-----ATGGTTGATGGTATCCCACTATTTTGAATCG 423
Qy 125 -----AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeu 139
Db 424 CGCTGGAATCTGGCGACGCACTTTCTGATATGGCACTTTTGAACGCGTAGAGTAGNG 483
Qy 140 ArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
Db 484 CGTGGCGGACAGGACTCATGACCGGCGGTAATCCATCTCGCGCAATTAATATGGTT 543
Qy 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArg 179

```

Db	1435	-----CGGTTGATACGCTGACTTACAGCATGGAGAAAAACACACACCGCCTTAC	1485
Qy	475	ProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeuSerGlyAsnTrpTyr	494
Db	1486	GCTGGTCTG-----GTGTTTGACATCAATGACAACTGG---	1518
Qy	495	PheThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThr	514
Db	1519	-----TCGACCTACGCCAGCTATACCTCTATTTTCCAGCCGCAA	1557
Qy	515	GlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLys	534
Db	1558	AATGATCGTCACAGTTTCAGGCAAAATATCTGGCTCCAATCACC-----	1599
Qy	535	HisLeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeu-----GlyTyrGluGly	552
Db	1600	-----GGTAACACTACGAGCTGGGTCTGAATCGGACTGGATGAAT	1641
Qy	553	AspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAla	572
Db	1642	AGCGTCTGACCAACCACGTTAGCCATCTCCGT-----	1674
Qy	573	GlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeu	592
Db	1675	-----ATTGAGCAGGATAATGTCGCTCAG---	1698
Qy	593	ValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLys	612
Db	1699	-----TCCACCGGTACACCTATCCCGCAGCAACGCGGAAACCGCCTATAAA	1746
Qy	613	ProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsn	632
Db	1747	-----GCGGTGATGGGACAGTCAGTAAAGGGGTGGAAATTGAA	1785
Qy	633	LeuProSer-----LeuProGlyArgGluAspAlaTyrGlyAsnArgProPheIleAla	650
Db	1786	CTCAACGGCGCAATTACCGACAACTGGCAGCTGACATTTGGCGGCAACGCGCTATATTGCA	1845
Qy	651	GlnAspAspGln-----AsnAlaProArg	658
Db	1846	GAGGATAACGAAGAAACGCCGTTAATCTCTATCTGCCACGC	1887

Search completed: November 17, 2003, 21:24:16
 Job time : 2666 secs

Db	544	CGAAACACGCGCAGCAGTCTGTGAATTTAAAGGCGATGTCGCGCGAAATACGCT---	597
Qy	180	LeuSerSerGlyAsnLeuGlyLys-----LeuThrSerGly	191
Db	598	-----AGCTGGAAACAAAGACCGTATGTCGCGGATTTACAAAGCCCACTCACCGAAGAC	651
Qy	192	GlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLys	211
Db	652	GGT-----AAAATCGCGCGCGCAATTGTCGCGCGCTAC---CAG	687
Qy	212	SerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArg	231
Db	688	AATAACGACTCATGGCTGACCGCTAC-----AACAGTGAAGAG	726
Qy	232	ArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPhe-----	248
Db	727	ACCTTCTCTCGGCAATGTCGATGCTGATTTAGCGCATCTTACGACGCTTTCAGCCGGT	786
Qy	249	-----TyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPro-----	263
Db	787	TACGAATATCAGCGCATGATGATTAATAGCCCTACCTGGGCGGTTTACCGCGCTGGAAT	846
Qy	264	-----AlaHisSerHisGluTyrAspAspCysHisAla-----AspIleIleTrp	278
Db	847	ACTGATGCGACGACGACAGTACGATCGCGACGCGATACCGACCTGACTGGCGGTAC	906
Qy	279	GlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTyrProHisLeu-----	294
Db	907	AACGTAAGAGATCAACAGGTCTTTATGACCTGACGAGCAGTGTGCTGATACCTGG	966
Qy	295	-----LeuThrGluAspValAspTyrAspAsnProGlyLeuSerCys	309
Db	967	CAAGCGACACTGATGCCACCCACTCTCGAAGTCGAATTTGACAGCAAAATGATGTATGTC	1026
Qy	310	GlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTrpIleAsp	329
Db	1027	-----GATCCCTATGTAACAAAGCGGATGGTATG-----	1056
Qy	330	LeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGlu	349
Db	1057	-----CTGGTTGGCCATACAGTAATTTATGACCTGGCTTGGT	1095
Qy	350	AlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLysAlaGlyAspAla	369
Db	1096	-----TATGTCGCGCGCACCGGTTGGAAACAGTGGCAAAAGTGAAGTGAATGCG	1143
Qy	370	ValGluAsnPhePhe-----	374
Db	1144	CTGGATTGTCGCTGACGCTAGTATGAATGTTGTTGCTGTCGACCAATCTAATGTTT	1203
Qy	375	-----AsnAsnGlnThrGlnAsnAlaAlaArgIleGluLeuArgHis	387
Db	1204	GGTGGCAGTTACAGCAAAACAAACAAATCGTTACTTCAGTTCATGGGCAACATCTTCCCG	1263
Qy	388	GlnProIleGlyArgLeu-----LysGlySerTrpGlyVal	399
Db	1264	GATGAATTTGGCAGTTTCTACAACTTAAATGGCAATTTCCCAAAACCGGCTGTCACCA	1323
Qy	400	GlnTyrLeuGlyGln-----LysSerSerAlaLeuSerAlaThrSerGlu	414
Db	1324	CAGAGCTGGCGGAGGAGGATACCAACATATGAATCGTTATATGCTGCCACTCGTGTC	1383
Qy	415	AlaValLysGlnProMetLeuLeuAspAsnLysValGlnHisTyrSerPheGlyVal	434
Db	1384	ACCTTCGCGATCGCTGCATCTG-----ATCCTCGGCGCA	1419
Qy	435	GluGlnAlaAsnTrpAspAsnPheThrLeuGluGlyValArgValGluLysGlnLys	454
Db	1420	CGTTATACCAACTGG-----	1434
Qy	455	AlaSerIleArgTyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeu	474

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 17, 2003, 17:07:47 ; Search time 2659 Seconds
(without alignments)
6928.467 Million cell updates/sec

Title: US-09-936-377-2
Perfect score: 4036
Sequence: 1 MAQTTLKPIVLSILLINTEL.....FLSDTPQMGSRFTGQVNVKF 758
Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Command line parameters:
-O=/cpn2_1/USPTO spooll/US09936377/runat_14112003_104450_22120/app_query.fasta_1.903
-DB=EST_QPMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09936377 @CQN 1 1 2135 @runat_14112003_104450_22120 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB	ID	Description
1	233.5	5.8	828	29	BZ565381	BZ565381 pacs2-164
2	185	4.6	983	29	BZ548740	BZ548740 pacs1-60
3	157	3.9	810	29	BZ551260	BZ551260 pacs1-60
4	153	3.8	960	29	BZ575443	BZ575443 msh2 4463
5	141	3.5	1218	29	BZ558672	BZ558672 p89801.3
6	130.5	3.2	1526	29	BZ576993	BZ576993 msh2 520
7	122	3.0	911	29	BZ554877	BZ554877 pacs1-60
8	120.5	3.0	1354	29	BZ554885	BZ554885 pacs1-60
9	119.5	3.0	1161	29	BZ552855	BZ552855 pacs1-60
10	118.5	2.9	2124	11	AY105818	AY105818 Zea mays
11	117	2.9	897	29	BZ572021	BZ572021 msh2 233
12	117	2.9	2280	11	AK005069	AK005069 Mus muscu
13	115.5	2.9	585	10	BG456354	BG456354 NF077H07P
14	114.5	2.8	765	13	BUI130232	BUI130232 603116821
15	112	2.8	610	28	BH374398	BH374398 AG-ND-153
16	112	2.8	1056	29	BZ548818	BZ548818 pacs1-60
17	112	2.8	1454	10	BF346294	BF346294 602018489
18	111.5	2.8	2341	11	AK031794	AK031794 Mus muscu
19	111	2.8	683	28	BH825061	BH825061 BACPP21-O
20	109.5	2.7	809	29	BZ597404	BZ597404 PUCCK35TD
21	109.5	2.7	1182	29	BZ565078	BZ565078 pacs2-164
22	109.5	2.7	1201	13	BX422599	BX422599 BX422599
23	109.5	2.7	1847	10	BF137846	BF137846 601782496
24	109	2.7	2864	11	AK043372	AK043372 Mus muscu
25	108.5	2.7	2169	11	AK016445	AK016445 Mus muscu
26	108	2.7	1120	13	BQ226095	BQ226095 AGENCOURT
27	108	2.7	5453	11	AK029845	AK029845 Mus muscu
28	107.5	2.7	837	13	BUI14889	BUI14889 603550605
29	107.5	2.7	960	13	BUI510302	BUI510302 AGENCOURT
30	107	2.7	262	10	BE123847	BE123847 BODAI10 BO
31	107	2.7	847	29	BZ548816	BZ548816 pacs1-60
32	107	2.7	879	29	BZ573965	BZ573965 msh2 3446
33	106.5	2.6	634	13	BUI065834	BUI065834 Fgr 8 K04
34	106.5	2.6	880	14	CA324304	CA324304 UI-M-FY0-
35	106	2.6	2531	11	AK016420	AK016420 Mus muscu
36	105.5	2.6	686	29	BZ564107	BZ564107 pacs2-164
37	105.5	2.6	1359	29	BZ563690	BZ563690 pacs2-164
38	105	2.6	858	10	BF969405	BF969405 602271677
39	105	2.6	858	13	BUI06765	BUI06765 603111672
40	104.5	2.6	932	9	AL969106	AL969106 AL969106
41	104.5	2.6	4015	11	AK036876	AK036876 Mus muscu
42	104	2.6	428	28	BH001150	BH001150 A2 Pirell
43	104	2.6	657	13	BUI064409	BUI064409 Fgr 4 M16
44	104	2.6	664	13	BUI060530	BUI060530 Fgr-C-1-E
45	104	2.6	696	13	BUI065008	BUI065008 Fgr_6_G07

ALIGNMENTS

RESULT 1
BZ565381
LOCUS
DEFINITION pacs2-164_522.sl pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION BZ565381
VERSION BZ565381.1
KEYWORDS GSS:
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 828)

AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. .828
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164-522"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 Whole genomic shotgun
library."

BASE COUNT 159 a 286 c 255 g 127 t 1 others

FEATURES

source
1. .828
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164-522"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 Whole genomic shotgun
library."

Alignment Scores:

Pred. No.: 1.56e-13 Length: 828
Score: 233.50 Matches: 77
Percent Similarity: 43.38% Conservative: 41
Best Local Similarity: 28.31% Mismatches: 114
Query Match: 5.79% Indels: 40
DB: 29 Gaps: 9

US-09-936-377-2 (1-758) x BZ565381 (1-828)

Qy 389 ProiledArgLeuLysGlySerTrpValGlnTyrLeuGlyGlnLysSerSerAla 408
Db 31 CCCTCGGACGCTGAGCGGGGTGTCGGCGGAGTTCGCCACACCGCTTCTCCGCC 90
Qy 409 LeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAenLysValGlnHis 428
Db 91 CTCGGC-----GAGGAAGCCTTCGTGCGCACACGAAACCGACAGG 132
Qy 429 TyrSerPheGlyValGlnAlaAsnTrp-----AspAnpHeThrLeuGlu 445
Db 133 GCGCGCTGTTCGCCCTGGAGGAA-----TGGAGCTCAGCGACCCCTCGACCTCAGC 186
Qy 446 GlyGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuIleAsp 465
Db 187 TTCGGCGCCGCTGGAG-----CACACCGCGTGGACCCCGCGCGCAAGGC 234
Qy 466 ArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArg 485
Db 235 AACGAGCGCTTCGCGGACGAC-----GGTTCGACAGCTTCACACCGCGC 282
Qy 486 SerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAla 505
Db 283 AGCTGTCCACCGCGCGGTGTACAGCTACGCGCGATCTGTCGTGCGCGCCACCTC 342
Qy 506 SerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAla 525
Db 343 AGCTACACCGAGCGCGCCCGACCTTCTACGAGCTGTACGCCACAGCTCGCGCGCGCC 402
Qy 526 ThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAnIleGlu 545
Db 403 ACCGGCAGCTACGAGGTAGCGGTCGCGACGCGACGAAGAAAGGCGGTCTCCACCGAC 462
Qy 546 LeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArg 565
Db 463 CTCGCCCTGCGCTTCGACACCGCGGTGCACAAAGGCGACGCGGTGGGTGTTCTACAGCGC 522
Qy 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLys----- 582

Db 523 TTCTCCAACATACATC-----GGGCTTCTCGCCAGCGGTGCATCGCAACGAGGAAGGC 576
Qy 583 -----SerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
Db 577 GAAGTGTGCGCGCCCGCATGACAGGGCGGTGCGGATACCTCTACAACGGCGTTCGC 636
Qy 600 AlaAspPheTyrGlyAlaGluGlyIleTyrPhe-----LysProThrPro 615
Db 637 GCGGAATTCACGCGTTCGAGCCAGGCGCATTCACCTTCTGGAAGCCCGTACGGCA 696
Qy 616 ArgTyrArgIleGlyValSerGlyAspTyrVal-----Arg 627
Db 697 ACTTTCGACTTGAACCTTTCGCGGGGATATACACCGCGCAAGAACACGAGGCAACGCGCG 756
Qy 628 GlyArgLeu-LysAsnLeuProSerLeuProGly 638
Db 757 AACGGCTTGAAGGATTGCCCGCGTTGCTGGA 790

RESULT 2

BZ548740/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1. .983

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"

/strain="1-60"

/db_xref="taxon:287"

/clone="pacs1-60_1393"

/clone_lib="pacs1-60"

/notes="clinical isolate 1-60 Whole genomic shotgun
library."

BASE COUNT 153 a 298 c 321 g 211 t

ORIGIN

Alignment Scores:

Pred. No.: 2.14e-08 Length: 983

Score: 185.00 Matches: 89

Percent Similarity: 39.23% Conservative: 53

Best Local Similarity: 24.59% Mismatches: 148

Query Match: 4.58% Indels: 75

DB: 29 Gaps: 10

US-09-936-377-2 (1-758) x BZ548740 (1-983)

Qy 64 AspThrLeuArgGlnLysAlaVal-----AsnLeuGlyAspAlaLeuAsp 78

Db 975 GATTTCCTTAACGCAAGGTTTACCTTGGAAAAAAGGAACCTCGGAA-ACCCTTAAA 917

Qy 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAlaProValIleArg 98

```

Db 916 GCTTTCCGGGGTGGC-TTAACTATTGGGAACGGGCCCAACCAACGGGTATCCGG 858
Qy 99 GylGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGly----- 115
Db 857 GGA-----TTGGAACGTTATCGGAATCCCGCTTTTCCAAA 822
Qy 116 -----AspMetAlaAspPheSerProAspHisAlaIleMetValAsp 129
Db 821 CGGGCTCGTGTCGTTGAACGCTTGTGGTTGCTTAGACAAAGGGTTGCCGGAAGAC 762
Qy 130 -ThrAlaLeuSerGlnGlnValGluLeuArgGlyProValThrLeuLeuTyr-SerS 149
Db 761 CCCAAAAGTTTCGAGCCCTCGAATTGTACGGCCGGCCCTTTGTTTACGGCTG 702
Qy 149 exGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProG 169
Db 701 GAAATGCAATCGCGGGGTGTTGAACAGCTCCGACCAACGATCCCAAGCGAC---CCGT 645
Qy 169 luAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeu 189
Db 644 TGGACGGCATCAACGAGCAGCGGGAATCGCTACGGCGGCGCGACACCCAGCCAGCC 585
Qy 189 hrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeu 209
Db 584 GCTCCGGCGCACTGGAGCGCGGCGACGCG---AACTTCGCCCTGACGTGACGCGCCCA 528
Qy 209 yrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspS 229
Db 527 GCGCGAGTTCACACGACGTC-----AGGATTCCCTGTGA 495
Qy 229 exProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPhe 249
Db 494 CGCCCATTCACGCGCGCGGCGAGATCGACGCGACAC---AAGC 448
Qy 249 yrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlu 269
Db 447 ATCGGTGCGAAGACAGCGCGCGCGGCGGACATCGGTGGCTCTATCACT 388
Qy 269 yrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeuG 289
Db 387 GGGAGCAGCGTTACGCGCGGCTCTCTCTAC----- 359
Qy 289 lnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerC 309
Db 358 -----AGCGGTACGACAGCAACTATGCTCGCGGCC----- 326
Qy 309 ysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysProTrpIleA 329
Db 325 -----GAGGACGAGTGGC-----CTGA 307
Qy 329 splLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPheG 349
Db 306 AGATGACGAGGACCGCTACGCTTCGCTCCGAGATCCGCGACCTCGAAGGCGCGGTCA 247
Qy 349 luAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspA 369
Db 246 CTTGCTGAGCTGGAGCGCGCTATACCAAGTACGAGTACGAGTACAGAAATCAGAGTGGCG 187
Qy 369 laValGluAsnPheAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnP 389
Db 186 AGACCGCGCACCATCTTCAAGAACGAGGCTACGAGGCGCGCATCGAGGCGCGCACCGCC 127
Qy 389 rolleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGlnLysSerSerAlaL 409
Db 126 CGTTCGCGCGCTGAAACGGGGTGGTGGCGCGGAGTTTCGCCAACAGCGCGCTTCTCGGCC 67
Qy 409 eu 409
Db 66 TC 65

```

RESULT 3
BZ551260

```

LOCUS      BZ551260      810 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION pacsl-60_3063.y2 pacsl-60 Pseudomonas aeruginosa genomic clone
ACCESSION  BZ551260
VERSION    BZ551260.1 GI:27154841
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 810)
AUTHORS    Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE       Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
JOURNAL     J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 20622216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES   Location/Qualifiers
            source          1..810
                        /organism="Pseudomonas aeruginosa"
                        /mol_type="genomic DNA"
                        /strain="1-60"
                        /db_xref="taxon:287"
                        /clone="pacsl-60_3063"
                        /clone_lib="pacsl-60"
                        /notes="clinical isolate 1-60 Whole genomic shotgun
                        library."
BASE COUNT 161 a 277 c 243 g 125 t 4 others
ORIGIN
Alignment Scores:
  Fred. No.: 1.29e-05      Length: 810
  Score: 157.00           Matches: 65
  Percent Similarity: 37.15% Conservative: 42
  Best Local Similarity: 22.57% Mismatches: 110
  Query Match: 3.89%      Indels: 72
  DB: 29                  Gaps: 10
US-09-936-377-2 (1-758) x BZ551260 (1-810)
Qy 130 ThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSer 149
Db 56 ACCCGGTGGCGCGCGCTCTAGAACTAGTAGTCCCGCC----- 94
Qy 150 GlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu 169
Db 95 -----GGTTTGTGAACAGCTTCGACAAACCGCATCCCGACGAA---CCCGTC 139
Qy 170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThr 189
Db 140 GACGGCATCCACGAGCGGCGGAACTGCGCTACGGCGCGCGCCACACACCGTAGCGCGC 199
Qy 190 SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyr 209
Db 200 TCCGGCGCACTGGAGCGCGCGGACGCG---AATTGCGCTGACGTGGAGCGCGCCAGC 256
Qy 210 ArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSer 229
Db 257 CGCGAGTTCAAACGACGTCAGGATTCGCGGTACGCCATCCCGCTCCAGCGCGC----- 304
Qy 230 ProArgArgPhe-----AlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArg 247
Db 305 CAGCGCAGATCGACCGCGCACCGCGCAAGCATCGGGTG----- 343
Qy 248 PheTyrArgArgThrTyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHis 267

```


library." 307 a 388 c 327 g 195 t 1 others

BASE COUNT 307 a 388 c 327 g 195 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 0.00106 Length: 1218

Score: 141.00 Matches: 82

Percent Similarity: 35.31% Conservative: 49

Best Local Similarity: 22.10% Mismatches: 139

Query Match: 3.49% Indels: 101

DB: 29 Gaps: 15

US-09-936-377-2 (1-758) x BZ558672 (1-1218)

Qy 384 GluLeuArgHisGlnProIleGlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGly 403

Db 72 GAGCACGGCACCAACCT---GGACACCTGGCG---GGA 104

Qy 404 GlnLysSer-SerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAs 423

Db 105 CAGCCCGCGGATGCCCTGGAAACGACGACGCGAC---TTTCCCGA 146

Qy 423 pAsnLysValGlnHisTyrSerPheGlyValGluGlnAlaAsnTrpAspAsnPhet 443

Db 147 TCCACCGTGAAGACTACGCCCTGTTCGCCAGACAGCATCAGCTGGAACGATGGAC 206

Qy 443 rLeuGluGlyValArgValGlu----- 451

Db 207 CTTCACTCCCGCGCTGGTTACGACTACACGCGACTGAGCGCGACATCACCAGCAGTT 266

Qy 452 -----LysGlnLysAlaSerIleArgTyrAspLysAlaLeuLleAspArgGl 467

Db 267 CCTCGCACCATGAAGCAGACGACACACCGCGGTGACGATCG---GA 314

Qy 467 uAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPh 487

Db 315 CAAGAAATGGACCGCGTTTCGCCCAAGTTCGCG-----GTCACTACGACTT 362

Qy 487 eAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHi 507

Db 363 CGCCACGACTACACTGGTATC-----GGCCAATACGCCAGCGGCTTC----- 405

Qy 507 sGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAs 527

Db 406 -----CGCACGCCACCGCCAGCGCTGTACGCTCGATTCGAGAACCTCGACGACG 458

Qy 527 nThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnLleGluLeuAl 547

Db 459 CTACACATCGAGCTAACCCACCTACCTACGCGGAAAGACGACGAGCTTCGACACGG 518

Qy 547 aLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGl 567

Db 519 GTTCGGCGCAAGTTCGACGACGAGCAGCTTCGGCGGTAGCGGTCTTACACAAATATCG 578

Qy 567 yAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGluAspAs 587

Db 579 CGACTTCATCGACGAAGACGCCCTGAT----- 606

Qy 587 pSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGl 607

Db 607 -----ACCGATAGCACCGCGCGACGCG 629

Qy 607 yGluIleTyrPheLysProThrProArgTyrArgIleGlyValSerGlyAspTyrValAr 627

Db 630 CCAGACC---TTCCAGTCCACACATCGAGCGCGGTGATCAAAGGCGTCGAGCTCAA 686

Qy 627 sGlyArgLeuLysAsn-LeuProSerLeuProGlyArgGluAspAlaTyrGlyAsnArg- 646

Db 687 GGGCGCGCTGAGCTGGCGGCTTTACGGCGCGCGCGGAGCTCCTACACCAAGGG 746

Qy 647 -----ProPhe-IleAlaGlnAspGlnAsnAlaPro 657

Db 747 CAACCGTTGGCTTAACGGCGCTAACGAGGTTTCGCAAAACAGCGGACCAACGCGGCCA 806

Qy 658 ArgValProAlaAlaArgLeuGlyPheHisLeuLysAlaSerLeuThrAspArgIleAsp 677

Db 807 GAAGCC-----GGATTCCAAACCAAGCGGTCTAAACCCCACTTTCAAC 851

Qy 678 AlaAsnLeuAspTyrTyrArgValPhe-----AlaGlnAsnLys 690

Db 852 CCGGGGAGATGTGTTTACGGCGCCCTTTTGGGGGTTAACGGAACGGAAGCCAGAAC--- 908

Qy 691 LeuAlaArgTyrGluThrArgThrProGly-----HisHisMetLeuAsnLeuGly 707

Db 909 -----CGGCAACCTTAACCGCGCGGGGCTTGCATTACGCGCTCGGGAACCTTGGG 962

Qy 708 AlaAsnTyrArgArgAsnThrArgTyr 716

Db 963 GTCAAAATACCCAAAGGCTATCGNCAT 989

RESULT 6

BZ576993 1526 bp DNA linear GSS 17-DEC-2003

LOCUS msh2_520.x1 msh Pseudomonas aeruginosa genomic clone msh2_520, genomic survey sequence.

DEFINITION BZ576993

ACCESSION BZ576993

VERSION BZ576993.1 GI:27212054

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

AUTHORS 1 (bases 1 to 1526)

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.

TITLE Whole-Genome-Sequence Variation among Multiple Isolates of Pseudomonas aeruginosa Library

JOURNAL J. Bacteriol., (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

FEATURES

source

1..1526

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"

/strain="M5H"

/db_xref="taxon:287"

/clone="msh2_520"

/notes="msh"

/notes="Environmental isolate. Whole genomic shotgun library."

BASE COUNT 320 a 461 c 383 g 362 t

ORIGIN

Alignment Scores:

Pred. No.: 0.018 Length: 1526

Score: 130.50 Matches: 47

Percent Similarity: 41.44% Conservative: 28

Best Local Similarity: 25.97% Mismatches: 75

Query Match: 3.23% Indels: 31

DB: 29 Gaps: 7

US-09-936-377-2 (1-758) x BZ576993 (1-1526)

Qy 75 AspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyAlaSerAla 94

Db 113 GACTTACCTGAAGACCATTCGCGCTTCGCGGTGATCCGACACGGCGGACGACGCGAC 172

Qy 95 ProValIleArgGlyGlnThrGlyArgGlyIleLysValLeuAsnHisGlyGluThr 114

Db 173 CCGGTGTGCGGGATGTTCGCTTCGCGCTTGAACATCTCCAAACGCGCGCATG--- 229

```

QY 115 GlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSer--- 133
Db 230 -----ATGCTCGTTCGACCGGATGACGCGCGACCTCTCAATCTCGCG 283
QY 134 -----GlnGlnValGluLeuLeuArgGlyProValThrLeuLeuTyrSerGly 150
Db 284 GAAACCTACCAAGCTCACCGTGATCAAGGCGCGCAGACGGTGCTCTGGGGCGCGGC 343
QY 151 AsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsn 170
Db 344 GCGTGGCGCGGAGCATC-----CTCTGAGCGTGAGCGGAGCGC 385
QY 171 GlyValSerGlyGluLeuGlyLeuArgLeu-----SerSerGly 183
Db 386 -----TTGCGGAACTCGCTCGCGGTCAACGCCAGCGCTGTCGCGCGCTCAATGCG 439
QY 184 AsnLeuGluLysLeuThrSerGlyGlyLeuAsnIleGlyLeuGlyLysAsnPheValLeu 203
Db 440 CGCTTCACAAAGTGTGGACGCGCGCGCGCAACCGGCTCGGC-----TACCTG 490
QY 204 HisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeu 223
Db 491 CGCTTACCGCAACACACGCGCAGTCCGACGATTACGAGACGCGCGCGGAATACCGTG 550
QY 224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
Db 551 -----CGTCGCGTGTGGAAGAGTGAACGCGCAGCGTTCGCGTGGG 592
QY 244 Trp 244
Db 593 TGG 595

RESULT 7
BZ554877 911 bp DNA linear GSS 17-DEC-2002
LOCUS
DEFINITION pacsl-60_4906.x1 pacsl-60 Pseudomonas aeruginosa genomic clone
ACCESSION BZ554877
VERSION BZ554877.1 GI:27162984
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 911)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source 1..911
location/Qualifiers
/mol_type="Pseudomonas aeruginosa"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_4906"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."
BASE COUNT 164 a 260 c 300 g 185 t 2 others
ORIGIN
Alignment Scores:

```

```

Pred. No.: 0.0651 Length: 911
Score: 122.00 Matches: 52
Percent Similarity: 36.67% Conservative: 36
Best Local Similarity: 21.67% Mismatches: 82
Query Match: 3.02% Indels: 70
DB: 29 Gaps: 9

US-09-936-377-2 (1-758) x BZ554877 (1-911)
QY 4 ThrThrLeuLysProIleValLeuSerIleLeuLeuIleAsnThrProLeuLeuAlaGln 23
Db 208 AGTTCGTTGTCGCGCCCTCGCGTCGCGATCGCTGGTGTGTCATCCATCGAGCGCGCAG 267
QY 24 AlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArg 43
Db 268 GCCGAGAGGAGCAAGCAAGAGAGCTGGGC-----ACGGTCACCGTGTGCGCGCATGGCTG 321
QY 44 ProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle---Ser 62
Db 322 GCGAAGCCGACCGAGCGGTGGTGCGAACAATCCCGTGGCGCAGCGTGGTGGTGGTGGT 381
QY 63 GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 382 CGGAAATGCTCGAGAGCGGTGGCGAGACGTCGCGACGTCGTGTCGCAAGTGCCTGGC 441
QY 83 IleHisAlaSerGlnTyrGlyGlyAlaSerAlaProVal-----Ile 97
Db 442 GTCAAGTGCAGGACACAAACCGGTACCGCGCGGCGACGACATCTCCCTGAAACGTGGGGTG 501
QY 98 ArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisGlyGluThrGlyAspMet 117
Db 502 CGTGGCTGACTTCACGC----- 519
QY 118 AlaAspPheSerProAspHisAlaIleMetValAsp----- 129
Db 520 -----CTGTGCGCGCGCTCGACGCTGATGATGACGCGCGCGCGCGCGCGCGCGCC 573
QY 130 -----ThrAlaLeuSer-----GlnGlnValGlu 137
Db 574 TACGCGCAGCGCAGCTGTGATGATGATCGCTGTCCATCGGCAACCTGGAGAGCACCAGC 633
QY 138 IleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAsp 157
Db 634 GTCTGTGCGCGCGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 693
QY 158 ValAlaAspGlyLysIleProGluLysMet-----ProGluAsn--- 170
Db 694 TTCTTGAACCGCGCGGATTCCCGAGAGGTTTTCGCGCAAAATCGGCGCACACCATCGAGCCTG 753
QY 171 ---GlyValSerGly----- 174
Db 754 CCGGTCAGCGCGGTGGAATAAGTTCAACCAAGGTTTCTTTGTGGGAACCGCGCGTCAA 813
QY 175 -----GluLeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGly 191
Db 814 ACGGCTGGCGTGGGCTTTTATTTCCGCGGGAAGGTGGCAATATATACGCGAGGCG 873

RESULT 8
BZ554885/c 1354 bp DNA linear GSS 17-DEC-2002
LOCUS
DEFINITION pacsl-60_491.s1 pacsl-60 Pseudomonas aeruginosa genomic clone
ACCESSION BZ554885
VERSION BZ554885.1 GI:27163009
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 1354)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of

```


JOURNAL
COMMENT

Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source

Location/Qualifiers

1..1354
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_491"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."

BASE COUNT 292 a 390 c 381 g 288 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 0.165 Length: 1354
Score: 120.50 Matches: 65
Percent Similarity: 36.10% Conservative: 35
Best Local Similarity: 23.47% Mismatches: 83
Query Match: 2.99% Indels: 94
DB: 29 Gaps: 14

US-09-936-377-2 (1-758) x BZ554885 (1-1354)

QY 39 ValGlyLysSerArgProAlaThrSerGlyLeuLeuHisThrSerThraAlaSerAsp 58
Db 715 GTTGGCGAAGCGGA-----CACAAAGTGTTGTGAATAAC 680

QY 59 Lysile-----lleSerGlyAspThrLeuArgGlnLys-AlaValas 72
Db 679 CATCTCCGGTCGCCCAACTGGTGGCTGCAGAAATGCTCAGAGCGGTGGCGAAA 620

QY 72 nLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSerGlnTyrglyGlyAl 92
Db 619 CGTCCGCGACGGTCTCGCAAGTGGCGCGCTGCAAGTGCAGGACAACACGGTACC 560

QY 92 aSerAlaProVal-----lleArgGlyGlnThrGlyArgAlleLysVa 107
Db 559 CGGACGCGACATCTCCTGAAAGTGGGGTGGTGGCTGACTTCACGC----- 511

QY 107 lleuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlalleMe 127
Db 510 -----CTGTCCGCGCGCTCGACGGTGAT 488

QY 127 tValasp-----ThrAl 131
Db 487 GATCGACGCGGTGCGCGCGCGGTGGCGCTTACGCGACGCGACGCTGTCGATGATGCC 428

QY 131 aleuSer-----GlnGlnValGluilleLeuArgGlyProValThrLeuTy 147
Db 427 GCTGTCCATCGGCAACTCGAGAGCATCATGCTGTGGCGCGCGCGCTGGTGGCGTA 368

QY 147 rSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysilleProGluLysMe 167
Db 367 CGGCCCCGCAAGCTCGGGGGGTGATCAACTCGTGACCCGGCGGATTCCGGAGAAGTT 308

QY 167 tProGluAsnGlyValSerGlyGluLeuGlyLeuargLeu-----SerSerGlyAs 184
Db 307 C-----TCCGGGGAATCGCGACCACTCGAGCATCGCGCGCGCGCGCG 263

QY 184 nLeuGlyLysLeuThr-----SerGlyGlylleAsnilleGlyLeuLysAs 200
Db 262 CTGAGAGAGCTCAACCGCGGTCTCTCGCGGGGACCGCGCGCGCGCGCGCGCGCG 208

QY 200 nPheValLeuHisThrGluGlyLeuTyArgLysSerGlyAspTyrrAlaValProArgTy 220
Db 207 -GTGGCGCTGTGTATTCGGGGTG-----AAGGGCGCGACTATC----- 169

QY 220 rArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAl 240
Db 168 -----CGGACGGCACACAGCAGCATGATAT 143

QY 240 aValLeuGlyTrpArgLysArgPheTyrrArgArgThrTyrrGerAspArgArgspGlnTy 260
Db 142 CGACGACGTGCTCTCAAGACCCATTGGCAGCTCCACGACGACGACCGACCG-- 94

QY 260 rGlyLeuProAlaHisSerHisGlyTyrrAspAspCysHisAlaAspIle 276
Db 93 ----TTGGCGGCAACTTCCACTACTAC---GACGCTACGCCGATATG 52

RESULT 9
BZ552855/c 1161 bp DNA linear GSS 17-DEC-2002
LOCUS pacsl-60_3853.y2 pacsl-60 Pseudomonas aeruginosa genomic clone
DEFINITION pacsl-60_3853, genomic survey sequence.
ACCESSION BZ552855
VERSION BZ552855.1 GI:27158033
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1161)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequencing variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source

Location/Qualifiers

1..1161
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_3853"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."

BASE COUNT 194 a 351 c 312 g 302 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 0.168 Length: 1161
Score: 119.50 Matches: 44
Percent Similarity: 37.06% Conservative: 29
Best Local Similarity: 22.34% Mismatches: 78
Query Match: 2.96% Indels: 46
DB: 29 Gaps: 5

US-09-936-377-2 (1-758) x BZ552855 (1-1161)

QY 218 ProArgTyrrArgAsnLeuLysArgLeuProAspSerPro-----ArgArgPhe 233
Db 623 CCGCGAGTTCAACGAAGCTGAGGATTCCCGCTACGCCACATTCCAGCGCGCGCGAG 564

QY 234 AlaAsnGlyGlnHisArgAlaValLeuGlyTrpArgLysArgPheTyrrArgArgThrTy 253
Db 563 ATCGAGCGCGACCGACGNCAGCATCGGGTGCAGACAGCGA----- 522


```

Qy 65 ThrLeuArg-----GlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db 298 GAGATGGCGCTGGCAACACCGCGATTAACTTGTCCGAATCACTGACCGGCGTGCCTGGT 357
Qy 83 Ile-----HisAlaSerGlnTyrGly-----89
Db 358 TTGAGGTACAAACCGCGCAGACTATCGCGAAGATTACAGTGTGCGATTCGGGATTT 417
Qy 90 GlyGlyAlaSerAlaProValIleArgGly-----GlnThrGlyArgArgIle 105
Db 418 GGCTCCGCTCCACTTACGGTATTCCGGGTATTGCGCTGTATGTGGACGGTATTCGCCGC 477
Qy 106 LysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAla 125
Db 478 ACCATGCCCGCGCGCAAGGCAACATCCAAC-----510
Qy 126 IleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeu 145
Db 511 -----ATCGATTAAAGCAGTGTGCAAAATGTGGAAGTGTGGTGGCCCTTCTCTGCC 564
Qy 146 LeuTyrSerSerGlyAsnValAlaGly 154
Db 565 CTGTAT-----GGCAACGGCTCTGGT 585

```

```

RESULT 14
BU130232/c 765 bp mRNA linear EST 25-NOV-2002
LOCUS 603116821F1 CSEQCHL21 Gallus gallus cDNA clone CHEST7216 5', mRNA
DEFINITION
sequence.
ACCESSION BU130232
VERSION BU130232.1 GI:25342143
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 765)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
2335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. .765
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST7216"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHL21"
/note="Organ: trunks; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; Modification of pBluescript
II KS(+) [Stratagene] vector to accommodate cDNA produced
with the T-trimmed protocol (Construction of
uni-directionally cloned cDNA libraries from messenger RNA
for improved 3' end DNA sequencing by Glenn Fu, et al.
U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
NotI and EcoRI. Ligate in double stranded adaptor
containing BsgI and BamHI sites
[5'ggcgcggtgcagccgcggtatccgaaaaaag]
[5'aattcttttttcggtatccggtgcagc]"

```

FEATURES

```

source
15
BH374398
LOCUS AG-ND-153H16.TF ND-TAM Anopheles gambiae genomic clone AG-ND-153H16
DEFINITION
610 bp DNA linear GSS 10-DRC-2001
BH374398

```

```

BASE COUNT 150 a 220 c 254 g 141 t
ORIGIN
Alignment Scores:
Pred. No.: 0.303 Length: 765
Score: 114.50 Matches: 61
Percent Similarity: 35.19% Conservative: 34
Best Local Similarity: 22.59% Mismatches: 88
Query Match: 2.84% Indels: 87
DB: 13 Gaps: 13
US-09-936-377-2 (1-758) x BU130232 (1-765)

```

```

Qy 15 LeuLeuAsnThrProLeuLeuAla-----GlnAlaHis 25
Db 660 CTCCTCAATACTCTTCTCATGCGCGCTTCTGTTAGGAATGTGAGAAGGGCATGTTTAC 601
Qy 26 GluThrGluGlnSerValGlnLeuGlyLeuValThrValValGlyLysSerArgProArg 45
Db 600 AGCACCGGGATGTGACCAAGATTCAGCCCTTGGTCCAGCTCAGTGCCTGAAACCGGCC 541
Qy 46 AlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGly-----63
Db 540 CGCAGGGCGGGAATCCACCACTCTTGGACCCACATCTTCATCATCTTCATCATCTCCTC 481
Qy 64 AspThrLeuArgGlnLysAlaValAsnLeuGly-----AspAlaLeuAspGlyValPro 81
Db 480 GAAGGTCTTTAGCAGGGTCTTGTCCAGCTTGGCTTAAAGACCGCTGCGCGGCTGGCT 421
Qy 82 GlyIleHis-----AlaSerGlnTyr 88
Db 420 GGGTCCCGCTCACGGGTGCGCTCTTCAACAGTCTTGAAGCCACCGTTGAGCAC 361
Qy 89 Gly-----GlyGlyAlaSerAlaProValIleArgGlyGlnThrGly 102
Db 360 GAGACCTCGCGGTGCCAAAGGCTCGGAACATCCACAGGCGCGGGGGCATAGAAGGT 301
Qy 103 ArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro 122
Db 300 GCCCAGCTCGTCCGCTCATACACAC---CACGTGGGTGCTGTTGCTACCCCGAGCCG 244
Qy 123 AspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyPro 142
Db 243 CCCAC---GTAGTCGGCAAGTGGCGCTCGTCGCGCAGCATGAATCATACGGGA---190
Qy 143 ValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLys 162
Db 189 -----GGACTTATCCCGGCACCTCTCGATGTCGAAGAAGGACGC 151
Qy 163 -----IleProGluLysMetProGluAsnGlyValSerGlyGluLeuGly 177
Db 150 GCCCGGAATGTCTCTCTTGAACCTCTCTGCGG-----GGC 115
Qy 178 LeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeu 197
Db 114 GTTGGCTCTTCGGGGG-----GTACCAAGAGGATCCAG-----79
Qy 198 GlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaVal 217
Db 78 -----CACCCGACGCCCGC-----64
Qy 218 ProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGlyGln 237
Db 63 -----CCCGACCTCGCGCGCCGCGCAACCGCTCGGACAGC 31
Qy 238 HisArgAlaValLeuGlyTrpArgLysArg 247
Db 30 CACTTGGCGCTGACCAAGCGCGCGGCCCAAA 1

```

```

, genomic survey sequence.
ACCESSION BH374398
VERSION BH374398.1 GI:17320540
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 610)
AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren
C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J.
and Collins, F.H.
TITLE Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)
MEDLINE 22542063
PUBMED 12655398
COMMENT Other-GSSs: AG-ND-153H16.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
FEATURES             Location/Qualifiers
     source            1..610
     /organism="Anopheles gambiae"
     /mol_type="genomic DNA"
     /strain="PEST"
     /db_xref="taxon:7165"
     /clone="AG-ND-153H16"
     /clone_lib="ND-TAM"
     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 215 a 101 c 110 g 184 t
ORIGIN
Alignment Scores:
Pred. No.: 0.398 Length: 610
Score: 112.00 Matches: 38
Percent Similarity: 42.07% Conservative: 31
Best Local Similarity: 23.17% Mismatches: 65
Query Match: 2.78% Indels: 30
DB: 28 Gaps: 5
US-09-936-377-2 (1-758) x BH374398 (1-610)
QY 5 ThrLeuLysProIleValLeuSerIleLeuLeuLeuAsnThrProLeuLeuAlaGlnAla 24
|||||
Db 119 ACAATGAAGAAGTTTTCAGTTTGGCATTATGCAGCCCAATATGTTTCGCAA 178
|||||
QY 25 HisGluThrGluGlnSerVal-----GlyLeuGluThrValThrValGlyLys 41
|||||
Db 179 GCGTTTGTAAGACAGACTCTCAAAGAAAATAGTATAGATGCCGTTGTAATGACAGACTC 238
|||||
QY 42 SerArg-----ProArgAlaThrSerGlyLeu 50
|||||
Db 239 TCTAAAACACATTCTGTAAGATAATCCATTACCTATCAAGAAGATTACGGCTAAATG 298
|||||
QY 51 LeuHisThrSerThrAlaSerAspLysIlelleSerGlyAspThrLeuArgGlnLysAla 70
|||||

```

```

Db 299 ATGGAG-----AAAACAATGAG 316
QY 71 ValAsnLeuGlyAspAlaLeu---AspGlyValProGlyIleHisAlaSerGlnTyGly 89
|||||
Db 317 ACAATATATATAGATGCATTGGTACAAATACTCCAGGACTGTAGCCGTAAACCGGG 376
QY 90 GlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsn 109
|||||
Db 377 CGGAATGATCAAAACCCCTTTATAAGAGGTTTAGGTATTAACAGAGTTCTTACACTG--- 433
QY 110 HisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAsp 129
|||||
Db 434 TATGATGGCCACAGACAGAGCGGACGATGGGAGATGAGCAGCATAGATAGTAGAC 493
QY 130 ThrAlaLeuSerGlnGlnValGluLeuLeuArgGlyProValThrLeuLeuTySerSer 149
|||||
Db 494 TCTTACACATTCCTCAGCGACAGTATCAAGAGACCATCCAGCTTAATGTACGATCC 553
QY 150 GlyAsnValAla 153
Db 554 GATGCTATAGCC 565
RESULT 16
LOCUS BZ548818/c
DEFINITION BZ548818
ACCESSION BZ548818
VERSION BZ548818.1 GI:27152399
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1056)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES             Location/Qualifiers
     source            1..1056
     /organism="Pseudomonas aeruginosa"
     /mol_type="genomic DNA"
     /strain="1-60"
     /db_xref="taxon:287"
     /clone="pacsl-60 1468"
     /clone_lib="pacsl-60"
     /note="clinical isolate 1-60 Whole genomic shotgun
     library."
BASE COUNT 204 a 356 c 320 g 176 t
ORIGIN
Alignment Scores:
Pred. No.: 0.878 Length: 1056
Score: 112.00 Matches: 48
Percent Similarity: 39.90% Conservative: 33
Best Local Similarity: 23.65% Mismatches: 77
Query Match: 2.78% Indels: 46
DB: 29 Gaps: 7
US-09-936-377-2 (1-758) x BZ548818 (1-1056)
QY 4 ThrThrLeuLysProIleValLeuSerIleLeuLeuLeuAsnThrProLeuLeuAlaGln 23
|||||

```



```

Db      555 AGTTCGTTGTCGGCCTGGCCGTCGCATCCGCTGGTGTGTCATGACGAGCGGCAG 496
Qy      24  AlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArg 43
Db      495 GCCGAAGAGGAGCAAGCAAGAGCTGGGC-----ACGGTCACCGTGTGTCGGCACTGGCTG 442
Qy      44  ProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle---Ser 62
Db      441 GCGAAGCCGACACAGCGGTGGTGGCAACCATCCCGTGGCGGACGCGTGGTGGTGGT 382
Qy      63  GlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGly 82
Db      381 CGGGAATGCTCAGAGCGGTGGCAGAACGTCGCGACGT-GTGGCAAGGTGCCCGC 323
Qy      83  IleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProVal-----Ile 97
Db      322 GTCCAAGTCGAGCAACAACAGGTACCGCGGCGAGCAGCATCTCCCTGAACGTGGGGTG 263
Qy      98  ArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisGlyGluThrGlyAspMet 117
Db      262 CGTGGCTGACTTCAGCC----- 245
Qy      118 AlaAspPheSerProAspHisAlaIleMetValAsp----- 129
Db      244 -----CTGTCGCGCGCTCGACGGTGATGATCGCGGCGGTGGCGGCC 191
Qy      130 -----ThrAlaLeuSer-----GlnGlnValGlu 137
Db      190 TACGGCCAGCGCAGCTGTCGATGATCCGCTGTCATCGCAACCTGGAGAGCATCGAC 131
Qy      138 IleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAsp 157
Db      130 GTGTCGCGCGCGCGCGCTCGGTGGCTACGGCCGAGAACGTCGCGGGGTGATCAAC 71
Qy      158 ValAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGly 177
Db      70  TTCGTGACCCGGCGATTCGGAGAGTTC---TCCGGCGGGGGATCTCTAGAGTCGAC 14
Qy      178 LeuArgLeu 180
Db      13  CTGAGGCTA 5

RESULT 17
BF346294
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Email: cgabs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LIA9423 row: a column: 23
High quality sequence stop: 62.
Location/Qualifiers
1. 1454
/organism="Homo sapiens"

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4154014"
/tissue_type="anaplastic oligodendroglioma with lp/19q loss"
/Tab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Brn67"
/notes="Organ: Brain; Vector: pCMV-SPORT6; Site 1: NCI; Site 2: Sall; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 609 a 502 c 305 g 38 t
ORIGIN
Alignment Scores:
Pred. No.: 1.39 Length: 1454
Score: 112.00 Matches: 88
Percent Similarity: 33.09% Conservative: 50
Best Local Similarity: 21.10% Mismatches: 168
Query Match: 2.78% Indels: 111
DB: 10 Gaps: 20

US-09-936-377-2 (1-758) x BF346294 (1-1454)
Qy 204 HisThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeu 223
Db 193 CATAAAGGAGCAAGAGAGCGAGCGAG-----ACAGAG 231
Qy 224 LysArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGly 243
Db 232 AAGCGACACCGGACACGACAGAG-----GGCGCATCACACGAGCGCGGCACTCAAC 285
Qy 244 TrpArgLysArgPheTyrArgArgThrTyrSerAspArg----- 256
Db 286 ACCAGGCAAGAGAGCAGCAGCAAAACAACCAACAAATAGCATAGAAAGAACCCCGAG 345
Qy 257 -----ArgAspGlnTyrGlyLeuProAlaHisSerHisGlu-----Tyr 269
Db 346 GAACGGGAAAGACACGCAACAAACACGACGAGCCACACACACGAGCGCGCAGGA 405
Qy 270 AspAspCysHisAlaAspIleIleTrpGlnLysSerLeuIleAsnLys-----ArgTyr 287
Db 406 AACGACGCGCGCAGATACTCACAACCGGCAACACAGAGCAACAAACAGAGACGCAC 465
Qy 288 LeuGlnLeuTyrProHisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeu 307
Db 466 CGCAAGAAAGCCGCCAC-----ACGAGGCCAACGACGACAAACAGAC----- 507
Qy 308 SerCysGlyPheHisAspAspAspAlaHisAlaHisAsnGlyLysPro--- 326
Db 508 ---GCCGGAACAAACAGAGAC-----GCACGCCACACGAGCCACCCCAA 549
Qy 327 -----TrpIleAspLeuArgAsnLysArg 334
Db 550 CAGACAAAGACACGCGCCAAAGACCAAGCCACCATGAGTGGACACACAGACGCGCAAGAGA 609
Qy 335 TyrGluLeuArgAlaGluTrpLysGlnProPheProGlyPheGluAlaLeuArgValHis 354
Db 610 -----ACACAAGCCGACACACGACGACCGCGACCC-----AGAACGAAACCCGCAC 654
Qy 355 LeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAspAlaValGluAsnPhePh 374
Db 655 CAGACGACAAACCGGAGCCAGCCCACTCCACCAACGCAAGCAAGAGACACTCAGAGCCCAACA 709
Qy 374 eAsnAsnGlnThrGlnAsnAlaArgIleGluLeuArgHisGlnProfile----- 390
Db 710 -AACAGCAACCGGAGCCAGCCACACACACACACAGAGAAAGAGACACTCAGAGCCCAACA 768
Qy 391 -----GlyArgLeuLysGlySerTrpGlyValGlnTyrLeuGlyGln----- 404
Db 769 CGACGGCACACAACTAGAGGAGCAAGACGCGGACACACACGCGCCAAACGAGCCCAAGCCCC 828
Qy 405 -----LysSe 406

```

```

Db      829 AGCCCGCCAGCAGACACAAAGACAAACACACACAGAGCGCGCACATACCGGC 888
      Qy      : rSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspHisLysVa 426
      Db      : ACGCGAAACAGGTAAGACAAAGAGCAAGAGAGCGCCCGGACAAACCAACCAACACAA 948
      Qy      : GlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGluG1 446
      Db      : ACGACACCGG-----GAACCGCACACAGGACAAACCCCGACAAACACACAG 993
      Qy      : yGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaLeuLeuAspAr 466
      Db      : ACGAAGCACACACAGGCGGACACACAGGACACACAAACAGAGAGGAGCGGACCGCGCG 1053
      Qy      : gGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGln--ThrAlaAr 485
      Db      : CGCAACACACACACACACAGGACAGCAGACACACACACACACACACACACACACACGCG 1113
      Qy      : gSerPheAlaLeuSerGlyAsnTrp---TyrPheThrProGln-----HisLysLeuSe 502
      Db      : AACCAGAGCGCGAAGAGGACACACAAACACACACACACACACACACACACAAACAAACA 1173
      Qy      : rLeuThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLy 522
      Db      : CCAGAACGAAAGCCACACACAGCACTCCCGACACAC-----CCCGACGCGACACACAA 1230
      Qy      : s-----HisValAlaThrAs 527
      Db      : ACGAACAGACACACAGCAGACACAGACACACACCGCCACCGACACACACACAGCAGGCG 1290
      Qy      : nThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
      Db      : AACAAACACACAAACAAACACCCACCAACCAACCAACCAACCAACCAACCAACCAACCA 1339

RESULT 18
AK031794
LOCUS
DEFINITION Mus musculus 11 days embryo head cDNA, RIKEN full-length enriched
library, clone:6230428F13 product:X-linked myotubular myopathy gene
1. full insert sequence.
AK031794 AK031794 2341 bp mRNA linear HTC 05-DEC-2002
VERSION AK031794.1 GI:26327616
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

```

```

20530913
MEDLINE
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
Saito,T., Okazaki,Y., Gojibori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staehli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustingich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
MEDLINE
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2341)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES
Location/Qualifiers
1. 2341
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6230428F13"
/db_xref="taxon:10090"
/clone="6230428F13"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="11 days embryo"
57. 1124
CDS

```

/note="unnamed protein product: X-linked myotubular
myopathy gene 1 (MGI:1099452, GBNM_019926, evidence:
BLASTN, 99%, match=1063)
putative"

/codon start=1
/protein_id="BAC27552.1"
/db_xref="GI:26327617"
/translation="MASASAKYNSHLENESIKYSDGVSDVSETVRLPGELLI
TEREYIICFNGPIKGRVYITNRLYRSLRSLDLSALIDVPLGVISRIKMGATSR
GENSYGLDITCKDLNRNLFALKQEGHSDRMFLVKHAPLAHNLPFLFVNEEKN
VDGWTVPVVEYRQGLPNHWRISFINKYELCETYPALLVPVPTSDDDRRIAT
FRSNRLPVLWSIHPENKVMVIRCSOPLVGMSGRKNDDEKYLVDIREINKOTSKLMI
YDARPSVNAVANKATGGYSDDAYQNSLSFLDINHVRWRESLKKVKDVIYVNTIEE
SHWLSLSLETHWLEHIKASIL"

BASE COUNT 760 a 389 c 457 g 735 t

ORIGIN

Alignment Scores:

Pred. No.: 3.12 Length: 2341
Score: 111.50 Matches: 100
Percent Similarity: 35.34% Conservative: 41
Best Local Similarity: 25.06% Mismatches: 149
Query Match: 2.76% Indels: 109
DB: 11 Gaps: 25

US-09-936-377-2 (1-758) x AK031794 (1-2341)

Qy 105 IleLysValLeuAenHisHisGlyGluThrClyAspMetAlaAaspPheSerProAspHis 124
Db 111 ATTAAAGAGTCTCAAGATGGAGTCAGTCAGATGTGTGAGACTGTCCTCGG--- 167
Qy 125 AlaMetValAspThrAlaLeuSerGlnGlnValGluLeuLeu----- 139
Db 168 -----CTCCAGGGGAGTACTTAATTAAGTAAAGAGATTTATACATATGTCCTTC 221
Qy 140 ArgGlyProValThrLeuLeuTySerSerGlyAenVal----- 152
Db 222 AATGGCCCATTT-----AAGGAGAGAGTTTACATACACAAATTATCGCTT 266
Qy 153 -----AlaGlyLeuValAspValAlaAaspGlyLysIle 163
Db 267 TATTAAAGAGTTTGGAAACGATCTGCTTAATCTGATGTTCTCTGGGTTGATA 326
Qy 164 Pro-----GluLysMetProGluAenGlyValSerGlyGlu-----LeuGlyLeuArg 179
Db 327 TCAAGATTGAAATAATGGGCGGCGCAAGTAGAGAGAGAAATTCATGCTTAGAT 386
Qy 180 LeuSerSerGlyAenLeuGluLysLeuThrSerGlyGlyLysAenLeuGlyLys 199
Db 387 ATTACTGTAAAGATTTCAGAAACCTG----- 413
Qy 200 AsnPheValLeuHisThrGluGlyLeuTyArgLysSer-----GlyAsp 214
Db 414 AGTTTGCATTGAAGCAAGAGGCGCACGACGAGAGAGATATGTTGATGCTTTGATAA 473
Qy 215 TyrAlaValProArgTyArgAsnLeuLysArgLeuProAspSerProArgArgPheAla 234
Db 474 CATGCTTCTCTCGCACACAATCTGCATTATT-----GCATTGTA 518
Qy 235 AsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyArgArgThrTySer 254
Db 519 AATGAAGAGAGATTAACTGGATGGATGGGTGG-----ACTGTTTATAATCCAGTTGAAGAA 572
Qy 255 AspArgArgAspGlnTyArgLeuProAlaHisSerHisGluTyArgAspCysHisAla 274
Db 573 TATAGAGGCGAG-----GGCTGCCCAATCACCAT----- 602
Qy 275 AspIleIleTrpGlnLysSerLeuIleAsnLysArgTy-----LeuGlnLeuTyPro 292
Db 603 -----TGGAGGATAAGTTTTTATTAAACAGTCTATGAGCTCTGTGACATACCT 653
Qy 293 HisLeuLeuThrGluGluAspValAspTyArgAsnProGlyLeuSerCysGlyPheHis 312

Db 654 GCTCTTTTGGTG-----GTTCCCTATCGGACC----- 680
Qy 313 AspAspAspAsp-----AlaHisAlaHisAlaHisAsnGlyLysPro----- 326
Db 681 TCAGATGATGATCTTAGAGAGATCGCAACGTTTAGATCCCGAAATCGCTTCTGTACTG 740
Qy 327 ---TrrPleAspLeuArgAsnLysArgTyGluLeuArgAlaGluTrrPlyGlnProphe 345
Db 741 TCGTGGATTCAACCAGAAAACAATGCTTATGCGCTGC-----AGTCACCTCTT 794
Qy 346 ProGlyPheGluAlaLeuArgValHisLeuAsnAspAsnAspTyHisHisAspGlyLys 365
Db 795 GTCGGTATGAGTGTGTAAGA-----AATAAGAT-----GACGAGAAA 833
Qy 366 AlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeu 385
Db 834 TACCTGGATGTGATCAGGAA-----ACTAACAAACAAACTTCTAAGCTCATGATTATGAT 890
Qy 386 ArgHisGlnProIleGlyArgLeu-----LysGlySerTrpGly----- 398
Db 891 GCAGACCCAGTGTAAATGAGTCGCCAACAGCAACAGGAGGAGGATATGAAGTAT 950
Qy 399 ValGlnTyrrLeuGlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGln 418
Db 951 GAGCATATCAAAACTCAGAACTTCTCTTACAGATTCATATATTCATGTTATGGA 1010
Qy 419 ProMetLeuLeuAspAsnLysValGlnHisTyrrSerPheGlyValGluGlnAlaAsn 438
Db 1011 GAATCTTTTA-----AAAAAGTGAAGATATTTTATCCCAACAGAGAATCTCAT 1064
Qy 439 TrpAspAsnPheThrLeuGluGlyGlyValArgValGluLysGlnLysAlaSerIle 457
Db 1065 TGGTGTGCC---AGTTGGAGTCTACTCATTTGGTTAGAACATATCAAGCAAGTATA 1118

RESULT 19
BH825061 683 bp DNA linear GSS 20-MAY-2002
LOCUS BACP21-012.y Pristionchus pacificus BAC ends Pristionchus
DEFINITION pacificus genomic, genomic survey sequence.
ACCESSION BH825061
VERSION BH825061.1 GI:21004090
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 683)
AUTHORS Srinivasan,J., Sinz,W., Lanz,C., Brand,A., Nandakumar,R., Raddatz
,G., Witte,H., Keller,H., Kipping,H., Pires da Silva,A., Jesse,T.,
Millare,J., de Both,M., Schuster,S.C. and Sommer R.J.
A BAC-based genetic linkage map of the nematode Pristionchus
pacificus
JOURNAL Unpublished
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES
source
1..683
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="var. California"
/db_xref="taxon:54126"
/clone_lib="Pristionchus pacificus BAC ends"
BASE COUNT 146 a 169 c 191 g 177 t
ORIGIN

Alignment Scores:

Pred. No.: 0.594 Length: 683

```

Score: 111.00      Matches: 45
Percent Similarity: 43.64%      Conservative: 27
Best Local Similarity: 27.27%      Mismatches: 61
Query Match: 2.75%      Indels: 32
DB: 28      Gaps: 8

US-09-936-377-2 (1-758) x BH825061 (1-683)

Qy 29 GlnSerValGlyLeuGluThrValThrValValGlyLys-----SerArgProArgAla 46
Db 45 CAGGATACCGCCCGGACTCTCGTGTACTGCTAAACCGTTTGAACAGCGCGGCAGC 104
Qy 47 ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleSerGlyAspThrLeu 66
Db 105 ACT-----GTGCTTGCCACCAACCAACCGTT-----GTGACCCGTCAGGATATCGAC 149
Qy 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyLysIleHisAlaSer 86
Db 150 CGCTGGTAGTCACCTCGGTCAATGATGTGCTGCCCGCTCTCCGGGCGTGCATATCAC 209
Qy 87 GlnTyrGlyGlyGlyAla-----SerAlaProValIleArgGlyGlnThrGlyArg 104
Db 210 CAACACCGCGGTTCAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 269
Qy 105 IleLysVal-----LeuAsnHisHisGlyGlyGluThrGlyAspMetAla 118
Db 270 GTGTGGTGTAAATTGATGGCGTACGCTGAAATCTGGCGGGGGTGCAGTGTCT---GCC 326
Qy 119 AspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIle 138
Db 327 GACCTTAGC-----CAGTTCCCTATTGGCTTGTCCAGCGTGTGAATAT 371
Qy 139 LeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspVal 158
Db 372 ATCCCGTGGCGCGCTCCGCTCTTTATGTTCCGATGCAATAGGCGGGGTGTGAATATC 431
Qy 159 AlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeu 178
Db 432 ATCAGACGCGC-----GATGAACCCCGAAGC 458

Qy 179 ArgLeuSerSerGly 183
Db 459 GAAATTCAGCAGG 473

RESULT 20
BZ697404/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1. (bases 1 to 809)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick
A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.
Maize Genomics Consortium
Unpublished
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: T7
Class: sheared ends.
Location/Qualifiers
1. 809
/organism="Zea mays"

FEATURES
source
BZ6565078
pac2-164_5068.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
pac2-164_5068, genomic survey sequence.

/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTal27F21"
/clone_lib="ZM 0.6 1.0 KB"
/note="vector: pCR4-TOPO; site_1: EcoRI; 0.6-1.0 kb high
Cot selected genomic DNA library"

BASE COUNT 111 a 286 c 250 g 162 t
ORIGIN
Alignment Scores:
Pred. No.: 1.09 Length: 809
Score: 109.50 Matches: 58
Percent Similarity: 34.78% Conservative: 22
Best Local Similarity: 25.22% Mismatches: 65
Query Match: 2.71% Indels: 85
DB: 29 Gaps: 10

US-09-936-377-2 (1-758) x BZ697404 (1-809)
Qy 471 LysGlnProLeuProAspLeuGlyAlaHisArgGlnThr----- 483
Db 796 AAGCAGGGGCTGCCGCGAACCGCGCGCAGCACGTGGTGGCGACATCAAGTTGT 737
Qy 484 AlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeu 503
Db 736 TCTCGTTCAGCGTCACGCGCGCGCGCGACATTCGCGCCCTAAGGACCCATTGAGCGAC 677
Qy 504 ThrAlaSerHisGlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHis 523
Db 676 CGAGTTCCCTCGGACCGCGACCCAGTCAGTCGGAGTACCCTGCTGAGTTAAGCAT 617
Qy 524 VallAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsn 543
Db 616 ATA-----ATAAGCGAGGAGAGAAAC 593
Qy 544 IleGluLeuAlaLeuGlyTyr-----GluGlyAspArgTrpGlnTyrAsnLeu 559
Db 592 TTACGA-----GGATTCCCTTAGTAACGCGGACGACCGGAGCAGCCGCTTG 542
Qy 560 AlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArg 579
Db 541 AGAATCGGGCGGCTCGCGCGCGGANTTGTATCTGGAGAGGCGTCTCAGCGACGCGAC 482
Qy 580 GlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyrAsnGlnSerGly 599
Db 481 GCGCCCAAGTTCTCTCGAAAGGACGCC----- 455
Qy 600 AlaAspPheTyrGlyAlaGluGlyGluIleTyrPheLysProThrPro----- 615
Db 454 -----TGGAGGGTGAGAGACCCCGCTCCGCGCGACCTGTGCGACCCAGC 410
Qy 616 -----ArgTyrArgIleGlyValSerGlyAspTyrVal----- 626
Db 409 AGGCCCGCTCAACGAGTCGGTGTGTTGGGAATGACGCCCAATCGGGCGGTAAACTCCG 350
Qy 627 -----ArgGly----- 628
Db 349 TCCAAGCTAAATACAGCGGAGAGACCCGATAGCAACAAAGTACCAGGGAAGATGAA 290
Qy 629 -----ArgLeuLysAsnLeuProSerLeuProGlyArgGlu----- 640
Db 289 AAGGACTTTGAAAGAGAGTCAAGAGTGTGCTTGAATTTGCCGGGGAAGCGGATGGGG 230
Qy 641 -----AspAlaTyrGlyAsnArgProPhe 648
Db 229 GCTGCGGACGCGCAC-----CGGCCGTAT 206

RESULT 21
BZ6565078/c
LOCUS
DEFINITION

```


Qy 150 GlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGlu 169
 Db 270 GGCCTCATTAAGGAAGAGTTTACATCAAAATATCTCTTTATTAAGAAGTTTGGA 329
 Qy 170 AsnGlyValSerGlyGluLeuGlyLeuArgLeuSer---SerGlyAsnLeuGlyLysLeu 188
 Db 330 ACGGATTCCTCTTAATATCTGATGTTCTCTCGGTGTGATCTCGAGAAATGAAAAATG 389
 Qy 189 -----ThrSerGlyGlyLeuAsn---IleGlyLeu----- 197
 Db 390 GGAGCGCGACAGTAGAGGAGAAATTCCTATGCTAGATATTACTTGTAAAGACATG 449
 Qy 198 ---GlyLysAsnPheValLeuHisThrGluGlyLeuTyraArgLysSer----- 212
 Db 450 AGAAACCTGAGGTTGCTTTGAAACAGGAAGCCACAGACAGAGATATGTTTGAGATC 509
 Qy 213 ---GlyAspTyraValAlaValProArgTyraArgAsnLeuLysArgLeuProAspSerProArg 231
 Db 510 CTCAGAGATACGGCTTTCCCTCGCTCACAGTCTGCCATATT----- 554
 Qy 232 ArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTyraArgLysArgPheTyraArg 251
 Db 555 GCATTTTAAATGAAGAAAGTTTAACTGGATGATGG-----ACAGTTTACAATCCA 608
 Qy 252 ThrTyraSerAspArgArgAspGlnTyraGlyLeuProAlaHisSerHisGluTyraAsp 271
 Db 609 GTGGAAGAAATACAGGAGCGAG-----GCCTTGCCCAATCACCAT----- 647
 Qy 272 CysHisAlaAspIleIleTyraGlnLysSerLeuLysAsnLysArgTyra-----LeuGln 289
 Db 648 -----TGGAATAAATCTTTTATTAATAAGTGTATGCTATGCTGTGTGAC 689
 Qy 290 LeuTyraProHisLeuLeuThrGluGluAspValAspTyraAspAsnProGlyLeuSerCys 309
 Db 690 ACTTACCTGCTCTTTGCTG-----GTTCCGTATCGT----- 722
 Qy 310 GlyPheHisAspAspAsp-----AlaHisAlaHisAlaHisAsnGlyLys 325
 Db 723 -----GCCTCAGATGATGACCTCCGAGAGTTGCACTTTTGTAGTCCGAAATCGAATT 776
 Qy 326 Pro-----TrpIleAspLeuArgAsnLysArgTyraGluLeuArgAlaGluTrpLys 342
 Db 777 CCAGTGCTGTATGATGATTCATCCAGAAATAGAGCGTCATGTGCGTTGC-----AGT 830
 Qy 343 GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyraHisHis 362
 Db 831 CAGCCTCTTGTGCGTATGATGAGTGGGAACGA-----ATAAAGAT----- 869
 Qy 363 AspGluLysAlaGlyAspAlaVal-----GluAsnPhe 373
 Db 870 GATGAGAAATATCTCGATGTTATCAGGGAGACTAATAACAAATTTTC 917
 RESULT 23
 LOCUS BF137846 1847 bp mRNA linear EST 24-OCT-2000
 DEFINITION 601782496F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4010734 5',
 mRNA sequence.
 ACCESSION BF137846
 VERSION BF137846.1 GI:10976886
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1847)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contract: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LAM9248 row: o column: 23

High quality sequence stop: 541.

FEATURES

source location/Qualifiers

1..1847

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CZECH II"

/db_xref="taxon:10090"

/clone="IMAGE:4010734"

/tissue="tumor, metastatic to mammary"

/lab_host="DH10B"

/clone_lib="NCI CGAP Lu30"

/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; transgenic model WNT-1, expression driven by
 MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
 dt. Library constructed by Life Technologies.
 Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 661 a 446 c 546 g 193 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 3.57 Length: 1847
 Score: 109.50 Matches: 121
 Percent Similarity: 31.73% Conservative: 50
 Best Local Similarity: 22.45% Mismatches: 191
 Query Match: 2.71% Indels: 179
 DB: 10 Gaps: 26

US-09-936-377-2 (1-758) x BF137846 (1-1847)

Qy 107 ValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerProAspHisAlaIle 126
 Db 519 GTAGNCAATGATCATGAAGAAGAT-----CTTGACTCGCGAAGAACTCATCGGC 569
 Qy 127 MetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeu 146
 Db 570 CAGGTGGAGAGCAAGTTGGCCAGCA-----CTG 599
 Qy 147 TyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLys 166
 Db 600 GTAGCAGTGTGCCACATGCTGGTTGGTTCCCTA-----AAA 638
 Qy 167 MetProGluAsnGlyValSerGlyGlu---LeuGlyLeuArgLeuSerSerGlyAsnLeu 185
 Db 639 CTACGGAAGGTTGAGTGGCGGAGACAGCGGTGCGAGACGCTCGCGGGGGGCAAC--- 695
 Qy 186 GluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThr 205
 Db 696 -----ACTGCAGTGTCTCAACCGCGGGGGGGGACAAAAT----- 731
 Qy 206 GluGlyLeuTyraArgLysSerGlyAspTyraValProArgTyraArgAsnLeuLysArg 225
 Db 731 ----- 731
 Qy 226 LeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTrpArg 245
 Db 732 ---GATCCACAGCCAGGCGG---ACCGGGGACACACCGTGGCACCTCAGCGGGCGT 785
 Qy 246 LysArgPheTyraArgArgThrTyraSerAspArgAspGlnTyraGlyLeuProAlaHis 265
 Db 786 AACCAAC-----CTTCGGTTTCC 803
 Qy 266 SerHisGluTyraAspAspCysHisAlaAspIleIleTrpGlnLysSerLeuLeuAsnLys 285
 Db 804 TCACACCATGATAGC-----CATCGG-----TGGTGGCTATCATGCTACGCAGA 848
 Qy 286 ArgTyraLeuGlnLeuTyraProHisLeuLeuThrGluGluAspValAspTyraAspAsnPro 305

849	CGCACCAAGACCAACAGCGCACATATA-----	875	
306	GlyLeuSerCysGlyPheHisAspAspAspAlaHisAlaHisAlaHisAsnGlyLys	325	
876	-----CGAAA-CACACACGAAGA	892	
326	ProTrpIleAspLeuArgAsnLys-----ArgTyrGluLeuArgAlaGluTrpLys	342	
893	CCAAGCATAGATCACAGAGAACAACACAGACTAAATATCGACAACAGACAGCCACAAA	952	
343	--GlnProPheProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis	362	
953	CACACCA-----CGAC	964	
362	isAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGln-----	379	
965	ACAACCGACGCAAAAGAAAACACACAGAACACACAGACCAACAACAGACCAACAAA	1024	
380	-----AsnAlaArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGly	396	
1025	CAAAACACCAACCAACGAGGGAG--CGAGGGACAGAGGGGAGGAACGCGGGA	1081	
396	erTrpGlyValGln-----TyrLeuGlyGlnLysSerS	407	
1082	GCGGAGCGCACAGAGAGAGGCCAAAGGAAGAGAGACAACGACGAGGAGAGAGCGAC	1141	
407	erAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeuAspAsnLysValG	427	
1142	AGGACGGGGACACCGCGCAAGAACCGGA-AGAAGAAAACGGAGAGCAGAA--	1197	
427	InHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPheThrLeuGlu	445	
1198	AGCATGAGAAGCACGAAGGGAGAGAGAGAGCGAAGAGAC-----ACAGAGCAGAGA	1251	
446	-----GlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLysAlaL	463	
1252	AGAGAACGGAGGAAGCGCGCAGAGCAGGACGACGACGACGAGAGAGAGCAGGA	1304	
463	euIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnT	483	
1305	-----GACCGAGGAGCGGGAAGAGAGAACCGAGAGAGAC-----CCAG	1344	
483	hrAlaArgSerPheAlaLeuSerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerL	503	
1345	CGCAGAGAGAGAGCGGACAGAGGAGAG-----	1373	
503	euThrAlaSerHisGlnGluArgLeuProSerThrThrGlnGluLeuTyrAlaHisGlyLysH	523	
1374	-----ACCCACGACAAAGAGAGAGAGAGACGACAGACGCGGAACCAACGACGAGGC	1425	
523	isValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsn-----	537	
1426	ACCAGCGCAGACAAGAGACGCA-----GGAACAAAGAGGAAGAGAGCGGAGCGGACAGACG	1482	
538	-----LysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyA	553	
1483	GCAAGAGAAACCCAGAGAGACGACAGACAGACAGACGACGACGAGGAGAAAGAGAGAG	1542	
553	spArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaG	573	
1543	GAACAGCGCAGCGGACGAGC-----AGCACACAGACCAACGACAGAGGAAGAGGCC	1596	
573	InThrLeuAsnAspGlyArgGlyProLysSerIleGlu-----AspAspSerGluMetL	591	
1597	AACAGGAAGGTGCACAAACACGACGAAAGACGAGAGAAACGACGACGAGAGCAACGAC	1656	
591	ysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAlaGluGlyGlu	608	
1657	AGCGGACGACGAGAAACAGACGCGGACGACAAACGCCAAAGAGAGAGGAGCGGAG	1709	
AK043372	2864 bp	linear	HTC 05-DEC-2002

RESULT 24
AK043372
LOCUS


```

US-09-936-377-2 (1-758) x AK016445 (1-2169)
QY 105 IleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaaspPheSerProAspHis 124
DB 250 AITAGAAAGTGTCTCAAGATGAGTCACTCAGGATGTGAGTGTGAGTGTCTCCCTCGG--- 306
QY 125 AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeu----- 139
DB 307 -----CTCCAGGGGAGTTACTAATTACTGAAAGAAAGTATTATTACATATATGCTCTTC 360
QY 140 ArgGlyProValThrLeuLeuTyrSerSerGlyAsnVal----- 152
DB 361 AATGGCCCCATT-----AAGGGAGAGATTATACATCAAAATTATCGTCTT 405
QY 153 -----AlaGlyLeuValAspValAlaAspGlyLysIle 163
DB 406 TATTTAAGAGTTTGAAGAGGATCTCTCTTAATCTTGAATCTCTCTGGGTGTGATA 465
QY 164 Pro-----GluLysMetProGluAsnGlyValSerGlyGlu-----LeuGlyLeuArg 179
DB 466 TCAAGAAATTGAAAAATGGAGGCGCACAAAGTAGAGGAGAAAAATTCCTATGCTTAGAT 525
QY 180 LeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLys 199
DB 526 ATTACTTTGAAGATTGAGAACCTG----- 552
QY 200 AsnPheValLeuHisThrGlyGlyLeuTyrArgLysSer-----GlyAsp 214
DB 553 AGGTTTGCAATTGAAGCAAGAGGCCACAGCAGAGAGATATGTTTGAGATCCTTTGTAAAA 612
QY 215 TyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgPheAla 234
DB 613 CATGCCCTTCTCTGCGCACACATCTGCCATTATTT-----GCATTTGTA 657
QY 235 AsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThrTyrSer 254
DB 658 AATGAAGAAGTTTAACGTGGATGGGTGG-----ACTGTTTATATCAGCTTGAGAA 711
QY 255 AspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspCysHisAla 274
DB 712 TATAGAAGCAG-----GSCCTGCCCAATCACCAT----- 741
QY 275 AspIleIleTrpGlnLysSerLeuIleAsnLysArgTyr-----LeuGlnLeuTyrPro 292
DB 742 -----TGGAGGATAAGTTTATTAAACAAGTCTATGAGCTCTGTGAGACATACCCT 792
QY 293 HisLeuLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSerCysGlyPheHis 312
DB 793 GCTCTTTTGGTG-----GTTCCCTATCGACC----- 819
QY 313 AspAspAspAsp-----AlaHisAlaHisAlaHisAsnGlyLysPro----- 326
DB 820 TCAGATGATGATCTTAGGAGGATCGCAACGTTTAGATCCGAAATCGGCTTCTCTGACTG 879
QY 327 ---TrpIleAspLeuArgAsnLysArgTyrGluLeuArgAlaGluTrpLysGlnProPhe 345
DB 880 TCGTGATTCACCCAGAAACAAATGCTCAITATGCGCTGC-----ACTGACCTCTT 933
QY 346 ProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisAspGluLys 365
DB 934 GTCGGTATGATGGTAAAGA-----AATAAGAT-----GACGAGAA 972
QY 366 AlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAlaArgIleGluLeu 385
DB 973 TACCTGGATGTGATCAGGAA-----ACTAACAAACAACTTCTAAGCTCATGATTATGAT 1029
QY 386 ArgHisGlnProIleGlyArgLeu-----LysGlySerTrpGly----- 398
DB 1030 GCACGACCCAGTGAATGCGTCGCAACAGGCAACAGGAGGAGATGATGAAGTGTAT 1089
QY 399 ValGlnTyrLeuGlyGlnLysSerAlaLeuSerAlaThrSerGluAlaValLysGln 418

```

```

DB 1090 GACGCATATCAAAACTGAGAACTTCCTCTTAGACATTCATAATATTCATGTTATGCGGA 1149
QY 419 ProMetLeuLeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsn 438
DB 1150 GAATCTTTA-----AAAAAGTGAAGATATTGTTATCCCAACATAGAGATCTCAT 1203
QY 439 Trp-----AspAsnPheThrLeuGluGly 446
DB 1204 TGGTGTCTCCAGTTTGAGTCTACTCATTTGTTAGAACATATCAAGCTGTGTTCTGACCGGT 1263
QY 447 GlyValArgValGluLysGln-----LysAlaSerIle----- 457
DB 1264 GCCATTCAAGTGGCAGACCAGTGTCTTCAGAAAGAGCTCGGTATCTTGTGCACTGCAGT 1323
QY 458 ---ArgTyrAspLys-----AlaLeuIleAspArgGluAsnTyr 469
DB 1324 GACGGATGGGACAGGACCCCTCAGCTGACATCCTTGGCCATCTGCTGATGTTGGACAGCTTC 1383
QY 470 TyrLysGlnProLeuProAspLeuGlyAlaHisArgGlnThrAlaArgSerPheAlaLeu 489
DB 1384 TACAGA-----ACTATTGAAGGCTTTGAGATA 1410
QY 490 -----SerGlyAsnTrpTyrPheThrProGlnHisLysLeuSerLeuThrAlaSerHis 507
DB 1411 TTGCTACAGAAAGAGTGG---ATAAGTTTTGGCCATAAAATTTGCATCTAGATAGTGCAT 1467
QY 508 GlnGluArgLeuProSerThrGlnGluLeuTyrAlaHisGlyLysHisValAlaThrAsn 527
DB 1468 GGTGATAAA----- 1476
QY 528 ThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnAsnIleGluLeuAla 547
DB 1477 -----AACCATGCTGATGCTGATCGATCT-----CCTATTTT 1509
QY 548 LeuGlyTyrGluGlyAspArgTyrGlnTyrAsnLeuAla----- 560
DB 1510 CTTTCAGTTTATGACTGTGTGTGGCAGATGTGAAACAGTTCGCCACAGCTTTTGAGTTC 1569
QY 561 -----LeuTyrArgAsnArgPheGlyAsnTyr 569
DB 1570 AATGAAGGCTTTTGTGATTACCGTTTGGATCATCTGTATAGCTGTGATGTTGTTGTTCTT 1629
QY 570 IleTyr 571
DB 1630 TTATTC 1635

```

```

RESULT 26
BO226095
LOCUS
DEFINITION
5', mRNA sequence.
ACCESSION
BO226095
VERSION
BO226095.1 GI:20407495
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1120)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML3346 row: i column: 24
High quality sequence stop: 470.

```

[illegible][illegible]


```

Db      2836 CGAGATAGAAATCCTAGCCATCAATACGACAGCTTGAAGGGAAGCCTCTGAGTGAAGC 2895
Qy      343 GlnProPhe---ProGlyPheGluAlaLeuArgValHisLeuAsnArgAsnAspTyrHis 361
Db      2896 CATCCACTTCTCCAGT-----GACAGAGAGAGCTGTCCAC 2931
Qy      362 HisAspGluLysAlaGlyAspAlaValGluAsnPhePheAsnAsnGlnThrGlnAsnAla 381
Db      2932 CTTGAAATTAAGAAACAGACAGATGGCTCACTTT-----TCAGATTC 2976
Qy      382 ArgIleGluLeuArgHisGlnProIleGlyArgLeuLysGlySerTyrGlyValGlnTyr 401
Db      2977 AGATACAAATACACACCT-ATCATTTGA-----GGATCCAA--- 3014
Qy      402 LeuGlyGlnLysSerAlaLeuSerAlaThrSerGluLysValGlnProMetLeu 421
Db      3015 -----AGCAAGA-AACAGCTGTCC-----CCAGTCCCAAGCCT----- 3049
Qy      422 LeuAspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTyrAspAsn 441
Db      3050 -----CGAGCCAGACGTACCCAGATGTGGCCTGAGTAAATGAAGACTGGGATCGA 3100
Qy      442 PheThrLeuGluGlyValArgValGluLysGlnLysAlaSerIleArgTyrAspLys 461
Db      3101 TCCACAGCCAGTGGCTTTAGGGCTTCT-----GACAGT 3136
Qy      462 AlaLeuIleAspArg---GluAsnTyrTyrLysGlnProLeuProAspLeuGlyAlaHis 480
Db      3137 GCAGATGCTGAACAAGAGAAACTTCTGTCTCAAGCATTTGGAGGACCTGGAGACCTGC 3196
Qy      481 ArgGlnThrAla-----ArgSerPheAlaLeuSerGlyAsnTyrPhe 495
Db      3197 GGCAGTGGGGATCTCGAGAGCTTGAGCAACAATCGTGGGAGT---ACTATG 3253
Qy      496 ThrProGlnHisLysLeuSerLeuThrAlaSerHisGlnGluArgLeuProSerThrGln 515
Db      3254 AGTTTGAATCATGAGGTCCAAATGGCTCGCAGTCAGCTGGGGCGACAGGCCAGCTTCCAG 3313
Qy      516 GluLeuTyrAlaHisGlyLysHisValAlaThrAsnThrPheGluValGlyAsnLysHis 535
Db      3314 GAACGGAGCAGTTTCACGGCCACACATATAGCCAAACAAT----- 3352
Qy      536 LeuAsnLysGluArgSerAsnAsnIleGluLeuAlaLeuGlyTyrGluGlyAspArgTyr 555
Db      3353 -----CGACAAACACCTGCTCCCTCAGACGTGGC----- 3392
Qy      556 GlnTyrAsnLeuAlaLeuTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeu 575
Db      3383 AGAAAGTCTGTAAACCTCGCGAAATGAAG-----CAAGAAATA 3421
Qy      576 AsnAspGlyArgGlyProLysSerIleGluAspAspSerGluMetLysLeuValArgTyr 595
Db      3422 AAGAGATCATGTCCCAACTCCGGTGGAG-----CTACACAGGTGACCTTATAC 3472
Qy      596 AsnGlnSerGlyAlaAspPheTyrGly-----AlaGluGlyGluIleTyrPheLys 612
Db      3473 AAGACTCTGGATGGAGACTTCGGTTCAGTGTGACACATGCGCTGCTG----- 3523
Qy      613 ProThrProArgTyrArgIleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsn 632
Db      3524 -----GAGAAAGCGCTGTATGTC-----AAAAAT 3547
Qy      633 Leu---ProSerLeuProGlyArgGluAspAlaTyrGlyAsnArgProphe 648
Db      3548 ATCCGNCAGCTGGGCAGGT-----GATTTGGGGCTTGAAGCCCTAC 3592

```

RESULT 28

BU414889/c

LOCUS

DEFINITION 60350605F1 CSEORBL06 Gallus gallus cdna clone CHEST52012 5', mRNA

ACCESSION BU414889

```

VERSION BU414889.1 GI:25907560
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
AUTHORS Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
MEDLINE Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED 22335534
COMMENT 12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
source
1..837
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST52012"
/sex="Male and female"
/tissue_type="Abdominal fat pad"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEORBL06"
/note="Vector: pBluescript II KS(+); Site 1: EcoRI;
Site 2: NotI; Modification of pBluescript II KS(+)
(Stratagene) vector to accommodate cDNA produced with the
T-trimmed protocol (Construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
,624). Cut pBluescript II KS(+) with NotI and EcoRI.
Ligate in double stranded adaptor containing BglI and
BamHI sites [5'ggccggcgagccgcggatccgaaaaaag]
[5'aattcttttttcggatccggggcgacgc]"
BASE COUNT 164 a 230 c 287 g 156 t
ORIGIN
Alignment Scores:
Pred. No.: 1.84 Length: 837
Score: 107.50 Matches: 60
Percent Similarity: 35.06% Conservative: 35
Best Local Similarity: 22.14% Mismatches: 91
Query Match: 2.66% Indels: 85
DB: 13 Gaps: 13
US-09-936-377-2 (1-758) x BU414889 (1-837)
Qy 11 LeuSerIleLeuLeuIleAsnThrProLeuLeuAla----- 22
Db 690 ATCTGCGTGAATCTCCTCAATACTCTTCTCATGCGCGCTTCTCTAGGAATGTGAGAA 631
Qy 23 ---GlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLys 41.
Db 630 GGGCATGTTTACAGACACACCGGAGTGCACGAGATTCAGGCCCTTGGTCCAGCTCAGTGCC 571
Qy 42 SerArgProArgAlaThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIle 61
Db 570 CTGNAACCGCCCGCCGAGGGGGGATCCACCACTCGGAACCTCTTGGACCCCACTTCTC 511
Qy 62 SerGly-----AspThrLeuArgGlnLysAlaValAsnLeuGly-----AspAlaLeu 77
Db 510 CATCATCTCTCGAAGGTCTTTAGCAGGGTCTTGTCTCCAGCTTGGCCTTAAAGACCGCTG 451

```

```

QY 78 AspGlyValProGlyIleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProValIle 97
Db 450 CGCGGGCTGGTGGCTCGCG-----CGTCACGGGGTGGCGCTCTCTTCACCCAGTTCTT 397
QY 98 ArgGlyGlnThrGlyArgArg-----104
Db 396 GAAGCCACCGTTCACACAGACACTCGGGTGGCCGAGAGGTCGGAAGTCGGAATCCACCGAGC 337
QY 105 -----11eIysValLeuAsnHisHisGlyGlyGly 115
Db 336 ACGGGGGGTAGAGGTGCCAGCTCGTCGCGGTATACACACCCAC-----GTGGGT 283
QY 116 AspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeuSerGlnGln 135
Db 282 GTCTGTGTGACCGCCCGCCCGC-----GTAGTCGCGCAAGTGGCGCTCGCTCGGAG 226
QY 136 ValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeu 155
Db 225 CATGAATCATACGGGA-----GGACTTATCCCGGCATCTC 190
QY 156 ValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGlu 175
Db 189 CTCGATGTC-----GAAGAAGACGCGCG-----166
QY 176 LeuGlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSer-----GlyGlyLeu 194
Db 165 -----GGAATGTGTCTCTCTGTAACCTCGCGGGGTGCGCTCTGCGGGGG-----115
QY 195 IleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAsp 214
Db 114 -----TACGAGGAGCATCCAGC 97
QY 215 TyrAlaValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAla 234
Db 96 ACCCGCAGCCG-----GCCCCGACCTGCGCGCTGACCGCGCGG 25
QY 235 AsnGlyGlnHisArgAlaValLeuGlyTyrArg 245
Db 57 TCGGACGACCTTGGCGCTGACCGCGCGG 25

RESULT 29
LOCUS BU510302/c
DEFINITION AGENCOURT_10116317 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:6504979 5', mRNA sequence.
ACCESSION BU510302
VERSION BU510302.1 GI:22816535
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.mcg.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14065 row: b column: 20
High quality sequence start: 10
High quality sequence stop: 664.
Location/Qualifiers
1..960
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
FEATURES
source

```

```

/clone="IMAGE:6504979"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 134"
/notes="Vector: PCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by Resgen, Invitrogen
Corp. Note: this is a NIH MGC Library."
BASE COUNT 207 a 301 c 226 g 226 t
ORIGIN
Alignment Scores:
Pred. No.: 2.24 Length: 960
Score: 107.50 Matches: 73
Percent Similarity: 31.69% Conservative: 36
Best Local Similarity: 21.22% Mismatches: 108
Query Match: 2.66% Indels: 127
Db: 13 Gaps: 14
US-09-936-377-2 (1-758) x BU510302 (1-960)
QY 79 GlyValProGlyIleHisAlaSerGlnTyrGlyGlyGlyAlaSerAlaProValIleArg 98
Db 929 GGGATCTCTGGTACAGAAAGCTGTTTCATGGGTAGGTTTTCAGCCAGATGACAGG 870
QY 99 Gly-----GlnThrGlyArgArg 104
Db 869 GGGCATAAGTGGCCAGAGGAAAGTGAATGCGAGGTAGGGTCCCGAGTTCGCGAAT 810
QY 105 IleLysValLeuAsnHisHisGlyGlyGlyThrGlyAspMetAlaAspPheSerProAspHis 124
Db 809 TTCTGTGATCAACATTCAGGGCCCCATCAATTCGAGGGAGCAGTGAAGAGACAC 750
QY 125 -----AlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeu 139
Db 749 AATCTGGCTAATAGCGGTTAAGTTGGTGTAGTGGCGCTCAATGTCGAGTTTCT 690
QY 140 ArgGlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAla 159
Db 689 ACGCAGATGTC-----ATAGATGGCTCATTTGTCACCAT 654
QY 160 AspGlyLysIleProGluLysMetProGluAsnGlyValSerGlyGlu-----Leu 176
Db 653 GAAGGCACAATC-----AGAGTGTCTCAGAGGTGTGGTGGTGGAGATGGAATTGTA 600
QY 177 GlyLeuArgLeuSerSerGlyAsnLeuGluLysLeuThrSerGlyGlyIleAsnIleGly 196
Db 599 GGGCTCAACCCACAGCAGTGGAAACCTGGGGGGCTGGGTAATGAGAACTCCAGCTTGA 540
QY 197 LeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyrArgLysSerGlyAspTyrAla 216
Db 539 CTTCTTTCC-----GTAATCCACAGAGAGCGCTCCATCAGCAGGAGGTCAAGCCAGA 486
QY 217 ValProArgTyrArgAsnLeuLysArgLeuProAspSerProArgArgPheAlaAsnGly 236
Db 485 GCC-----AGTTCCTCCCGCCAAAGAGCTGTGGAAC 456
QY 237 GlnHis-----ArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThr 252
Db 455 CAAGAAGCCCTGGAGACCCGTCGCTGTCAGCCAGCTTGGCAATCTGTCAGGACAG 396
QY 253 TyrSerAsp-----255
Db 395 GTCAATGATCTCTTGGCAATGGTGTAGTGGCCAGCGCATAGTTATTGGCAGCATCTC 336
QY 256 -----ArgArgAspGlnTyrGlyLeuProAlaHisSerHisGlu 268
Db 335 CTTCCTGTGATGAGCTGCTCAGGATGAGAGCTGGCGGTAGGTGCCAGCTCGCACTTC 276
QY 269 TyrAspAspCysHisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeu 288
Db 275 ATCGATGAC-----CGTGGGTTC 258

```

```
QY 289 GlnLeuTyrProHis-----LeuLeuThrGluAsp 299
|||:|||||
Db 257 CAGGCTACGACACTGCCCGGCACATGCTTGCACGCTCTCTCTCACTGAAGAGGT 198
|||:|||||
QY 300 ValAsp-----
|||:|||||
Db 197 GTTGAAGGAGTCACTCTCTCCCAATGGTCTTGTCATCTGGCATCAGGCTG 138
|||:|||||
QY 302 -----TyrAspAsnProGlyLeuSerCysGlyPheHisasp-----
|||:|||||
Db 137 GATGCCATGTTCCAGGAGTAGATCCACAGCGCATTCGCCGATCTGGACACAGCGCTG 78
|||:|||||
QY 314 -----AspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIleAspLeu 330
|||:|||||
Db 77 GCCAAGGTGGATGAGATGCATCAGCGATAGCAACGG-----TTAGATCCC 27
|||:|||||
QY 331 ArgAsnLysArg 334
|||:|||||
Db 26 AGGAATTCCGG 15

RESULT 30
BE123847/c
LOCUS BE123847 262 bp mRNA linear EST 14-JUN-2000
DEFINITION BOD410 BOD-24 Bos taurus cDNA, mRNA sequence.
ACCESSION BE123847
VERSION BE123847.1 GI:8517161
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 262)
AUTHORS Guo,Z. and Zhang,Y.
TITLE Differential expressing genes from bovine cumulus oocyte complexes
before and after cultured in vitro
JOURNAL Unpublished
COMMENT Contact: Zekun Guo
Laboratory of Developmental Biology
Northwestern Sci-Tech University of Agriculture and Forestry
22 Xinong Road, Yangling, Shaanxi province, 712100, P.R.China
Tel: 86-029-7098576
Fax: 86-029-7092176
Email: devbio@wau.edu.cn; guozekun@163.net.

FEATURES
source
1..262
Location/Qualifiers
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/sex="female"
/tissue_type="ovarium"
/cell_type="cumulus oocyte complex"
/dev_stage="culture in vitro for 24 hour"
/clone_lib="BOD-24"
/notes="Sequences obtained by DDRT-PCR method"

BASE COUNT 88 a 49 c 41 g 84 t
ORIGIN
Alignment Scores:
Pred. No.: 0.397 Length: 262
Score: 107.00 Matches: 26
Percent Similarity: 52.22% Conservative: 21
Best Local Similarity: 28.89% Mismatches: 35
Query Match: 2.65% Indels: 8
DB: 10 Gaps: 2

US-09-936-377-2 (1-758) x BE123847 (1-262)

QY 526 ThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgSerAsnIleGlu 545
|||:|||||
Db 259 ACCAATCGCTATGAAGTGGGATGCCAATTTGAACACAGACAAATTTTCAACGGAT 200
|||:|||||
QY 546 LeuAlaLeuGlyTyrGluGlyAspArgTrpGlnTyrAsnLeuAlaLeuTyrArgAsnArg 565
|||:|||||
```

```
Db 199 TTGATTGGATACAAACAGATCAGCTGGAGTTTTTTTGCAATGGATTCTATATATCAC 140
|||:|||||
QY 566 PheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyProLysSerIleGlu 585
|||:|||||
Db 139 ATTACTGATTTTATTTATATCAATCCTACCGAGAA-----TCAATTGAT 95
|||:|||||
QY 586 AspAspSerGluMetLysLeuValArgTyrAsnGlnSerGlyAlaAspPheTyrGlyAla 605
|||:|||||
Db 94 GACAAT-----GAAGTTTTTAATATGTTCAAAATAATGCGGCTCTTTTGGAGGT 44
|||:|||||
QY 606 GluGlyGluIleTyrPheLysProThrPro 615
|||:|||||
Db 43 GAAATAGGCTTGCAATTTCCACCTCATCCG 14
|||:|||||

RESULT 31
BZ548816/c
LOCUS BZ548816 847 bp DNA linear GSS 17-DEC-2002
DEFINITION pacsl-60_1466.sl pacsl-60 Pseudomonas aeruginosa genomic clone
ACCESSION BZ548816
VERSION BZ548816.1 GI:27152397
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 847)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1..847
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_1466"
/clone_lib="pacsl-60"
/notes="clinical isolate 1-60 whole genomic shotgun
library."

BASE COUNT 129 a 274 c 274 g 169 t
ORIGIN
Alignment Scores:
Pred. No.: 2.11 Length: 847
Score: 107.00 Matches: 41
Percent Similarity: 44.19% Conservative: 16
Best Local Similarity: 31.78% Mismatches: 52
Query Match: 2.65% Indels: 20
DB: 29 Gaps: 4

US-09-936-377-2 (1-758) x BZ548816 (1-847)

QY 68 GlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIle---HisAlaSer 86
|||:|||||
Db 414 CAGCGCCCGCGCAACCTTGACGCGCTGGCAACATCATCAGCGCATCACCAGCCCAAC 355
|||:|||||
QY 87 GlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgIleLys 106
|||:|||||
Db 354 ACCCTCGCGGTATCCCGAGGACGCGGTGATGAAGCGCGC----- 316
|||:|||||
QY 107 ValLeuAsnHisHisGlyGluThrGlyAsp-----MetAlaAspPheSerProAsp 123
|||:|||||
```

```

Db      315 -----TTCCGGCGCAACCGCGACGGCTCGATCATCGCGCGGCGATGCGCTCG 268
Qy      124 HisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIleLeuArgGlyProVal 143
Db      267 GTGCAGGCGCCGAATTTCCACGCCACCCCGCCAGCCAGCTCGAGGTCTGAAGGGGCCACA 208
Qy      144 ThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLysIle 163
Db      207 TCCTGTCTATGGCATCCAGACCCGGCGGGGGTGGTCAACGTGGTG----- 160
Qy      164 ProGluLysMetProGluAsnGlyValSerGlyGluLeuGlyLeuArgLeuSerSerGly 183
Db      159 ---AGCAAGAGCCGCAACTGCAACAGCCCAACGCCCTGACCTCGTGGCTCGGCTAC 103
Qy      184 AsnLeuGluLysLeuThrSerGlyGly 192
Db      102 GCCACAGGGCGCAACGGCAGCGGGCGC 76

```

```

RESULT 32
LOCUS   BZ573965
DEFINITION msh2_3446.x1 msh Pseudomonas aeruginosa genomic clone msh2_3446,
ACCESSION BZ573965
VERSION   BZ573965.1 GI:27209026
KEYWORDS  GSS.
SOURCE    Pseudomonas aeruginosa
ORGANISM  Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 874)
AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
          Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol., (2002) In press
COMMENT   Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu
          Class: shotgun.
FEATURES
     source
     1..874
     /organism="Pseudomonas aeruginosa"
     /mol_type="genomic DNA"
     /strain="MSH"
     /db_xref="caxon:287"
     /clone="msh2_3446"
     /note="Environmental isolate. Whole genomic shotgun
     library."

```

```

BASE COUNT 175 a 278 c 277 g 143 t 1 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      2.2      Length:      874
Score:          107.00    Matches:      67
Percent Similarity: 35.38%  Conservative: 25
Best Local Similarity: 25.77%  Mismatches: 101
Query Match:      2.65%    Indels:      67
DB:               29      Gaps:      14

```

```

US-09-936-377-2 (1-758) x BZ573965 (1-874)

```

```

Qy      30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArg----- 45
Db      97 TCCGCGCGGCTGTGTGGCTGCTCGGACCGCCAGAAACGTGAACGCTCGCTGGAGGAC 156
Qy      46 ---AlaThrSerGlyLeuLeuHisThrSerThr-----AlaSerAspLysIle 60

```

```

Db      157 ACGCAGTCTCTCGGTCTCGCTCACCGCGCACGCGACATTGACCCCAAGCAGAC----- 210
Qy      61 lIleSerGlyAspThrLeuArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyVal 80
Db      211 -----CGGCAACGCTCGGTGTCGCGAGGTGATCAATGGCAGTCCCAA 252
Qy      81 ProGlyIleHisAlaSerGlnTyrGlyGlyAlaSer-----AlaProValIleArg 98
Db      253 CGTGTCTACAC-----CGACTCGTGGCGCGCGGATCATCCGCGCGCAGGACACCCA 306
Qy      99 GlyGlnThrGlyArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAsp----- 116
Db      307 GGGCCCGCAACACGGCGCAGAACTGTTCTGGGGCGGCACAGTGC CGCGCGCAGCATCAA 366
Qy      117 -----MetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu 132
Db      367 CTGATGGCCACTACTGTAACACAGAGAT---GTTCTTCGGCGCGACCTCGGTCTG 423
Qy      133 Ser---GlnGln-ValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAs 151
Db      424 GGAGCTCGACAGCATCGAGGTGTTCCGCGCGCGCGCAGACACCTCCAGGGCGCCAAACGC 483
Qy      151 nValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGl 171
Db      484 CATGCCCGGGCGGATC---ATCGTCAACACCAAGGACCCGACCTTCAGCCCCGAGGCTGG 540
Qy      171 yValSerGlyGluLeuGlyLeuArgLeuSer-----SerGlyAs 184
Db      541 CTACCAAGGGGGGATGTCAGCTACCATTCGCGCGCGAGATTGATCGGAAGTCCGGCCC 600
Qy      184 nLeuGluLysLeuThrSerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHi 204
Db      601 GTTGGCAAGACTTTGGCGCGCGCC-----GGCGGGGGG----- 634
Qy      204 sThrGluGlyLeuTyrArgLysSerGlyAspTyrAlaValProArgTyrArgAsnLeuLy 224
Db      635 -ACTACGCGGTGCGCACACCTTCATTGATACGCAACCCCGAAGTTCACGACAGGGG 693
Qy      224 sArgLeuProAspSerProArgArgPheAlaAsnGlyGlnHisArgAlaValLeuGlyTr 244
Db      694 GCACCGACCGAGGACTTCG-----GGCCCTTAAGCGCGCGGCCAAGCTG 738
Qy      244 pArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPro 263
Db      739 G-----TTGGCTTGCCA 751

```

```

RESULT 33
LOCUS   BU065834/c
DEFINITION Fgr 8_K04_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
ACCESSION BU065834
VERSION   BU065834.1 GI:22506123
KEYWORDS  EST.
SOURCE    Gibberella zeae
ORGANISM  Gibberella zeae
REFERENCE 1 (bases 1 to 634)
AUTHORS   Trail,P., Xu,J.-R., San Miguel,P., Halgren,R.G. and Kistler,H.C.
TITLE     Analysis of expressed sequence tags from Gibberella zeae (anamorph
          Fusarium graminearum)
JOURNAL   Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE   22508120
PUBMED    12620255
COMMENT   Contact: Frances Trail
          Department of Plant Biology
          Michigan State University
          East Lansing, MI 48824, USA
          Tel: 517 432 2939
          Fax: 517 353 1926
          Email: trail@msu.edu

```



```

Plate: 8 row: K column: 04.
FEATURES source
  Location/Qualifiers
    1..634
      /organism="Gibberella zeae"
      /mol_type="mRNA"
      /strain="NRRL 31084"
      /db_xref="taxon:5518"
      /clone_lib="Nitrogen-starved mycelia"
      /note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 128 a 231 c 154 g 121 t
ORIGIN
Alignment Scores:
Pred. No.: 1.56 Length: 634
Score: 106.50 Matches: 67
Percent Similarity: 38.70% Conservatives: 22
Best Local Similarity: 29.13% Mismatches: 78
Query Match: 2.64% Indels: 63
DB: 13 Gaps: 13
US-09-936-377-2 (1-758) x BU055834 (1-634)
Qy 18 ThrProLeuLeuAlaGlnAlaHisGluThrGluGlnSerValGlyLeuGluThrValThr 37
Db 587 ACTCTTTACAGCAAGGGCAGCAAGAGCAGCCATGCGAGGCGCACCGACCCCTTCAG 528
Qy 38 ValVal-----GlyLysSerArgProArgAla 46
Db 527 GCCAGCAGCACCGTTGGCTGCTCGGAGCCCTTGTAGGGTGGCAGCGGGGGTGTCT 468
Qy 47 ThrSerGlyLeuLeuHisThrSerThrAlaSerAspLysIleIleSerGlyAspThrLeu 66
Db 467 GGTGGCGGGCTCTTACAGCGGTGGTGGGAATGATGGGACGACGAGTGGTGGGCGAGC 408
Qy 67 ArgGlnLysAlaValAsnLeuGlyAspAlaLeuAspGlyValProGlyIleHisAlaSer 86
Db 407 GCGGTGGTCTCAACACCTCGCGCTGGTGGTGCACACGACCAAGTGGAGTGGGCGAGC 348
Qy 87 GlnTyrGlyGlyAlaSerAlaProValIleArgGlyGlnThrGlyArgArgIle---- 105
Db 347 -----GGCGGTGGTGGTGGCTTCTCTCTCGCGGTGGACTTGGCTCAGACTCCTT 294
Qy 106 -----LysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPhe 120
Db 293 GGGGGGTTCTGCGAGGTTCTC-----GTTAGCGGGAAGGACCTGTTGATGGC 243
Qy 121 SerProAspHisAlaIleMetValAspThr----- 130
Db 242 AAGTCGGTACCACTCGTCAATGACACAGAGGTAGCTTGGATCGGACAGCGCTCAA 183
Qy 131 AlaLeuSerGlnGlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGly 150
Db 182 GCCTTCGACACACAGCGGAGTC-----GGTCTTGTCA----- 147
Qy 151 AsnValAlaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsn 170
Db 146 -----CTGGTCTC-----GCTGGGATGGCTTGTGCGGCAAG-----AAT 108
Qy 171 GlyValSerGlyGluLeuGlyLeuArgLeuSer---SerGlyAsnLeuLeuGluLysThr 189
Db 107 GCGCACTTGGGACGTCGCGAAGAGACTGAGCTGAGCGGCAAC----- 63
Qy 190 SerGlyGlyIleAsnIleGlyLeuGlyLysAsnPheValLeuHisThrGluGlyLeuTyr 209
Db 62 AGCGCGGAGACCAAGCAAGCGGGA-----GTACTTCATTTTGAATGTGTGTTT 12
Qy 210 ArgLysSerGlyAspTyrAlaValProArg 219
Db 11 -----GTTCTCTGT 3
CA324304 880 bp mRNA linear EST 26-NOV-2002

```

DEFINITION UI-M-FY0-ccp-e-01-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
IMAGE: 6822602 5', mRNA sequence.
CA324304
VERSION CA324304.1 GI:24542402
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 880)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1..880
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone IMAGE: 6822602"
/tissue type="whole brain"
/dev stage="embryo 13.5, 14.5, 16.5, 17.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FY0"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoRI; Site_2: NotI; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a NotI site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoRI adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the NotI site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 242 a 230 c 235 g 168 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 2.51 Length: 880
Score: 106.50 Matches: 57
Percent Similarity: 32.96% Conservatives: 32
Best Local Similarity: 21.11% Mismatches: 108
Query Match: 2.64% Indels: 73
DB: 14 Gaps: 9

US-09-936-377-2 (1-758) x CA324304 (1-880)

Qy 102 GlyArgArgIleLysValLeuAsnHisGlyGluThrGlyAspMetAlaAspPheSer 121
Db 145 GGGCGCCAGTCACGACCAAGCAGTCCACCCGCCCTCTGAGTCTCACAGTCA 204
Qy 122 ProAspHisAlaIleMetValAspThrAlaLeuSerGlnGlnValGluIle---LeuArg 140
Db 205 CCAGGCCGAGAGGTGATGACTATGCTCAGCCCGCAGCGCATGGTAGAGTGACCTTCAT 264
Qy 141 GlyProValThrLeuLeuTyrSerSerGlyAsnValAlaGlyLeuValAspValAla--- 159

RESULT 34
CA324304
LOCUS


```

QY      521  lylyshisValAlaThrAsnThrPheGluValGlyAsnLysHisLeuAsnLysGluArgS 541
Db      1607  GGCCACACTATAGCAACAACT-----CGCA 1633

QY      541  erAsnAsnIleGluLeuAlaLeuGlyTyrgluGlyAspArgTrpGlnTyrAsnLeuAlaL 561
Db      1634  GCAACACCCCTGCCTCAGAGCTGGC-----AGAAAGTCGTAAACC 1675

QY      561  euTyrArgAsnArgPheGlyAsnTyrIleTyrAlaGlnThrLeuAsnAspGlyArgGlyP 581
Db      1676  TGGGAAATGAAG-----CAAGAAATAAAGAGATCATGTCCC 1714

QY      581  rlyysSerIleGluAspAspSerGluMetIysLeuValArgTyrAsnGlnSerGlyAlaA 601
Db      1715  CAACCTCGGTGGAG-----CTACAAAGGTGACCTTATACAGGACTCTGGCATGG 1765

QY      601  spPheTyrGly-----AlaGluGlyGluIleTyrPheLysProThrProArgTyrA 618
Db      1766  AGGACTTCGGGTTCCAGTGTGCAGATGGCTGCTG----- 1800

QY      618  rglleGlyValSerGlyAspTyrValArgGlyArgLeuLysAsnLeu---ProSerLeuP 637
Db      1801  -----GAGAAAGGGCGTGTATGTC-----AAAAATATCCGCCAGCTGGGC 1840

QY      637  roGlyArgGluAspAlaTyrGlyAsnArgProPhe 648
Db      1841  CAGGT-----GATGTTGGGGCTTGAAGCCCTAC 1869

RESULT 36
BZ564107
LOCUS   686 bp DNA linear GSS 17-DEC-2002
DEFINITION
pac2-164_4573.y2 pac2-164 Pseudomonas aeruginosa genomic clone
ACCESSION
BZ564107.1 GI:27189244
VERSION
BZ564107.1
KEYWORDS
GSS.
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 686)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
CONTACT: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Source
1. 686
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164_4573"
/clone_lib="pac2-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."
BASE COUNT 127 a 227 c 229 g 103 t
ORIGIN

Alignment Scores:
Pred. No.: 2.22 Length: 686
Score: 105.50 Matches: 31
Percent Similarity: 47.62% Conservative: 9
Best Local Similarity: 36.90% Indels: 3
Query Match: 2.61% Gaps: 3
DB: 29
US-09-936-377-2 (1-758) x BZ563690 (1-1359)

```

```

Query Match: 2.61% Indels: 5
Db: 29 Gaps: 3
US-09-936-377-2 (1-758) x BZ564107 (1-686)

QY      678  AlahenLeuAspTyrTyrArgValPheAlaGlnAsnLysLeuAlaArgTyrGluThrArg 697
Db      144  GCGCGGTTCGAGTCGACACACCGCCCTCGCAGCACCGCGTGGCGGAGGAAGAACTCTCC 203

QY      698  ThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGly 717
Db      204  ACCGACGGCTAC-----ACCACCTCGGCCCGCCCTCGGCTACACTTCGACCTCGGC 257

QY      718  GluTyrAsnTyr-----TyrVallyAlaAspAsnLeuLeuAsnGlnSerVallyAla 735
Db      258  GAGACCCCTGGTGGCTTCGTCAAGGGCACCACCACTGACCAACGACGCTCGCTAC 317

QY      736  HisSerSerPheLeuSerAsp---ThrProGlnMetGlyArgSerPheThrGlyGlyVal 754
Db      318  GCCAGTTTCGATCTCGCGCGCGGCTGCGCGCGGACGCGGACGCGCATCGAGGCGGGGTG 377

QY      755  AsnVallyysPhe 758
Db      378  AAGGTGGCGGTTCC 389

RESULT 37
BZ563690
LOCUS   1359 bp DNA linear GSS 17-DEC-2002
DEFINITION
pac2-164_4378.y2 pac2-164 Pseudomonas aeruginosa genomic clone
ACCESSION
BZ563690
VERSION
BZ563690.1 GI:27188088
KEYWORDS
GSS.
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 1359)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
CONTACT: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Source
1. 1359
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164_4378"
/clone_lib="pac2-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."
BASE COUNT 271 a 385 c 278 g 422 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 5.96 Length: 1359
Score: 105.50 Matches: 31
Percent Similarity: 47.62% Conservative: 9
Best Local Similarity: 36.90% Mismatches: 39
Query Match: 2.61% Indels: 5
DB: 29 Gaps: 3
US-09-936-377-2 (1-758) x BZ563690 (1-1359)

```

Qy 678 AlaasnLeuAspTyrTyArgValPheAlaGlnAsnLysLeuAlaAArgTyrGluThrArg 697
 Db 147 GCGGGTGCAGCTGGAACACGCCCGCTCGCAGCACCGCGTCCGCGGAGAACTCTCC 206
 Qy 698 ThrProGlyHisHisMetLeuAsnLeuGlyAlaAsnTyrArgArgAsnThrArgTyrGly 717
 Db 207 ACCGACGGCTAC-----ACCACCTCGCGCCACCTGGGCTACAACTTCGACCTCGC 260
 Qy 718 GluTrpAsnTrp-----TyrValLysAlaAspAsnLeuLeuAsnGlnSerValTyrAla 735
 Db 261 GAGAGCGCTCGCTGGCTCGTCAAGGGCCACCACTGACCAACACGACCGTGGCTAC 320
 Qy 736 HisSerSerPheLeuSerAsp---ThrProGlnMetGlyArgSerPheThrGlyVal 754
 Db 321 GCCAGTTCGATCTTCGCGACCGGTGCGGACGGGACCGGATCGAGCGGGGGTG 380
 Qy 755 AsnValLysPhe 758
 Db 381 AAGGTGGCGTTC 392

RESULT 38

BF969405 858 bp mRNA linear EST 22-JAN-2001
 LOCUS 602271677F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4359666 5',
 mRNA sequence.

ACCESSION BF969405

VERSION BF969405.1 GI:12336620

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgepbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LHAM9999 row: j column: 19

High quality sequence stop: 734.

Location/Qualifiers

FEATURES

1..858
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4359666"
 /tissue_type="adrenal cortex carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_84"
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site 1:
 NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT
 primed. Average insert size 1.229 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH MGC Library."
 BASE COUNT 185 a 250 c 289 g 134 t
 ORIGIN

Alignment Scores:

Pred. No.: 3.46 Length: 858
 Score: 105.00 Matches: 72
 Percent Similarity: 35.71% Conservative: 43
 Best Local Similarity: 22.36% Mismatches: 94
 Query Match: 2.60% Indels: 113
 DB: 10 Gaps: 18

US-09-936-377-2 (1-758) x BF969405 (1-858)

Qy 30 SerValGlyLeuGluThrValThrValValGlyLysSerArgProArgAlaThrSerGly 49
 Db 40 GCAGTCGGATATGAACACATCGCCCTTCGGGTGTCGAGCGGAGG----- 87
 Qy 50 LeuLeuHisThrSerThrAlaSerAspLysLeleSerGlyAspThrLeuArgGlnLys 69
 Db 88 -----CAAGCCCAAGTGCAGTAGAGTACAAAGGGGAGAC-----CAAGAC 129
 Qy 70 AlaValAsnLeuGlyAspAlaLeuAspGly----- 79
 Db 130 CTTCTTCCAGAGAGATATCTCATGCTCTCAGAGATGAAGGAGATCGCGAAGC 189
 Qy 80 ValProGlyIleHisAlaSerGlnTyrGlyGlyAla----- 92
 Db 190 CTACCTGGGGGCAAGGTGCACAGCGGTCTATACGGTCCCGGCTATTTCAACGACTC 249
 Qy 93 SerAlaProValIleArgGlyGlnThrGlyArgArgIleLysValLeuAsnHisGly 112
 Db 250 GCAGCGCCA-----GGCCACCAGAGCGCAGG-----CACCATCAGCGG 288
 Qy 113 GluThrGlyAspMetAlaAspPheSerProAspHisAlaIleMetValAspThrAlaLeu 132
 Db 288 ----- 288
 Qy 133 SerGln-GlnValGluIleLeuArgGlyProValThrLeuLeuTyrSerSerGlyAsnVa 152
 Db 289 GCTCAATGTGTGGCATCATCAAGAGCCCAAGCGCGCGCCATCGCTACGGC----- 343
 Qy 152 LalaGlyLeuValAspValAlaAspGlyLysIleProGluLysMetProGluAsnGlyVa 172
 Db 344 -----CTGACACAGAGAGGTG 360
 Qy 172 LserGly-----GluLeuGly----- 177
 Db 361 CGCGGGCGCGAGAGAACGTCTCATCTTTGACTGGCGGTGGCATTTTCGAGGTGC 420
 Qy 178 ----LeuArgLeuSerSerGlyAsnLeuGlu-----LysLeuThrSerGlyGlyIleAsnI 195
 Db 421 CATCTGACCATCGAGGTGCATCTTCGAGGTGAAGTCCACGCGCGCGATACCCACT 480
 Qy 195 eGlyLeuGlyLysAsnPe-----ValLeuHis---ThrGluGlyLeuTyrAr 210
 Db 481 GGGC---GGTGAGGACTTCGACCAACCGCATGTGTGAGCCACCTGGCGGAGAGTTCAAGCG 537
 Qy 210 GLYS-----SerGlyAspTyrAlaValProArgTyrArgAsnLeuLy 224
 Db 538 CAAGCACAGAGAGACATTGGGCCCAACAGCGCGCTGAGCGGTGGCCACCGTTCG 597
 Qy 224 sArgLeuProAspSerProArgArgPheAlaAsn---GlyGlnHisArgAlaValLeuG 243
 Db 598 GAGCGCGCAAGCGCACCTGAGCTCTGCCACGAGCGAGCATCGATCGACTCGC 657
 Qy 243 YTrpArgLysArgPheTyrArgArgThrTyrSerAspArgAspGlnTyrGlyLeuPr 263
 Db 658 TCTACAGGGCGGTGGACTTCTTATACGTCCATCAGCGCGCGCTTCGAGGAGCTCAATGCC 717
 Qy 263 oAlaHisSerHisGluTyrAspAsp-----CysHi 273
 Db 718 GACTCTTTCGGGACCTGGAGCGGTGGAGAGCGCTGGCCAGCGCAGCTGGGCAAG 777
 Qy 273 sAlaAsp-----IleIleTrpGlnLysSerLeuIleAsnLysArgTyrLeuGlnLeuTy 291
 Db 778 GGCGGATCCACAGATCGGTGGGGCGGTGCCCTTTTCCAAAGTCCCACGCTGGGATTTT 837
 Qy 291 rPro 292
 Db 838 CCGG 841

RESULT 39

BU106765/c

LOCUS

BU106765

858 bp

mRNA

linear

EST 25-NOV-2002

```

DEFINITION 603111672F1 CSEQCHL12 Gallus gallus cdna clone CHEST59018 5', mRNA
sequence.
ACCESSION BU106765
VERSION BU106765.1 GI:25309076
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE 1. (bases 1 to 858)
AUTHORS Boardman,P., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
MEDLINE Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED 22335534
COMMENT 12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
FEATURES
     source
     1..858
         /organism="Gallus gallus"
         /mol_type="mRNA"
         /strain="White Leghorn, HiseX"
         /db_xref="taxon:9031"
         /clone="CHEST59018"
         /dev_stage="36"
         /lab_host="DH10B"
         /clone_lib="CSEQCHL12"
         /note="Organ: heart; Vector: pBluescript II KS(+); Site 1:
         EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)
         [Stratagene] vector to accommodate cDNA produced with the
         7-trimmed protocol (Construction of uni-directionally
         cloned cDNA libraries from messenger RNA for improved 3'
         end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
         ,624). Cut pBluescript II KS(+) with NotI and EcoRI.
         Ligate in double stranded adaptor containing BspI and
         BamHI sites [5'ggcgcgtgcagccggatccgaaaaag]
         [5'aattcttttcggtccgggtgcacgc]"
BASE COUNT 164 a 242 c 288 g 164 t
ORIGIN
Alignment Scores:
Pred. No.: 3.46 Length: 858
Score: 105.00 Matches: 44
Percent Similarity: 38.89% Conservative: 26
Best Local Similarity: 24.44% Mismatches: 68
Query Match: 2.60% Indels: 42
DB: 13 Gaps: 8

US-09-936-377-2 (1-758) x BU106765 (1-858)
QY 15 LeuileAsnThrProLeuLeuAla-----GlnAlaHis 25
Db 687 CFCCTCAATACCTCTTCATCGCCGCTTCCTGTTAGGATGTTGAGAGGGCATGTTAC 628

QY 26 GluThrGluGlnSerValGlyLeuGluThrValThrValValGlyLysSerArgProArg 45
Db 627 AGCACACGGGAGTGTGACCAAGATTCAGGCCCTTGTCCAGCTCAGTGCCTCGAAACCGGCC 568

QY 46 AlaThrSerGlyLeuLeuHisThrThrAlaSerAspLysIleIleSerGly----- 63
Db 567 CCACGGCGGGGAAATCCACCACCTGGAACCTTCTTGGACCCCAATTCATCATCTCCTC 508

QY 64 AspThrLeuArgGlnLysAlaValAsnLeuGly-----AspAlaLeuAspGlyValPro 81
Db 507 GAAGGCTTTAGCAGGGTCTTCTCCAGCTTGGCCCTTAAGACCGCTCGCGGGGTGGCT 448

```

```

QY 82 GlyIleHis-----GlyGlyAlaSerAlaProValIleArgGlyGlnThrGly 88
Db 447 GGGCTCCGCCGTCACAGGGGTGCCCTCTCTCCACCGAGTCTTGAAGCCACCGTTCAGCAC 388

QY 89 Gly-----GlyGlyAlaSerAlaProValIleArgGlyGlnThrGly 102
Db 387 GGAGACCTCGCGTGCCTCCAAAGGCTCGGAACATCCACAGGCGCGGGGGCATAGAGGT 328

QY 103 ArgArgIleLysValLeuAsnHisHisGlyGluThrGlyAspMetAlaAspPheSerPro 122
Db 327 GCCACGCTCGTCCGCTCATACACACAC---CACGTGGGTGTGCTGCTGACCCCGGCG 271

QY 123 AspHisAlaIleMetValAspThrAlaLeuSerGlnValGluLeuLeuArgGlyPro 142
Db 270 CCCAC---GTAGTCGGCAAGTGGCCCTCGTCGCAGCATGAATCATACGGGA--- 217

QY 143 ValThrLeuLeuTy-SerSerGlyAsnValAlaGlyLeuValAspValAlaAspGlyLys 162
Db 216 -----GGACTTATCCCGGCACCTCTCTGATGTCGAAGAAGGACGC 178

RESULT 40
LOCUS AL969106
DEFINITION AL969106 XGC-gastrula Silurana tropicalis cdna clone TGasi27p16 5',
mRNA sequence.
ACCESSION AL969106
VERSION AL969106.1 GI:25792701
KEYWORDS EST.
SOURCE Silurana tropicalis (western clawed frog)
ORGANISM Silurana tropicalis
REFERENCE 1. (bases 1 to 932)
AUTHORS Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2002
JOURNAL Unpublished
COMMENT Contact: Taylor R
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TGasi27p16.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
FEATURES
     Location/Qualifiers
     1..932
         /organism="Silurana tropicalis"
         /mol_type="mRNA"
         /db_xref="taxon:8364"
         /clone="TGasi27p16"
         /dev_stage="gastrula (stages 10.5-13 mixed)"
         /lab_host="Escherichia coli XL1-blue"
         /clone_lib="XGC-gastrula"
         /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
         was oligo dT primed from 5ug of poly A+ RNA from stages
         10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
         into pCS107 with EcoRI at the 5' end and NotI at the 3'
         end."
BASE COUNT 271 a 185 c 242 g 233 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 4.39 Length: 932
Score: 104.50 Matches: 80
Percent Similarity: 38.18% Conservative: 46
Best Local Similarity: 24.24% Mismatches: 117
Query Match: 2.59% Indels: 88
DB: 9 Gaps: 16

US-09-936-377-2 (1-758) x AL969106 (1-932)

```

```
Qy 233 PheAlaAsnGlyGlnHisArgAlaValLeuGlyTyrArgLysArgPheTyrArgArgThr 252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
16 TTTCCTAAGGACCAATGTAGA-----TTTCTTCTTCTACC 48
Qy 253 TyrSerAspArgArgAspGlnTyrGlyLeuProAlaHisSerHisGluTyrAspAspCys 272
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 TTGGCAGATTGTATTTCCTGTTGGT-----GACAATAAT 84
Qy 273 HisAlaAspIleIleTyrGlnLysSerLeuIleAsnLysArgTyrLeu----- 288
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 AATGCCATTCCTTAAGTTTGAACAGCCAGATGCTGGATCCTTATCTCATAAAGGAATG 144
Qy 289 GlnLeuTyrProHisLeuThrGluGluAspValAspTyrAspAsnProGlyLeuSer 308
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
145 GATATATATGGTACTTCTGCTGCGCTGAAGTGCAGTGGAGATGTGGAATAATCTTGGC 204
Qy 309 CysGlyPheHisAspAspAlaHisAlaHisAlaHisAsnGlyLysProTrpIle 328
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
205 TGCCGCTTGTTCAAATATATCTGATCAGCATGCT-----GAGCCATGGGTG 249
Qy 329 AspLeuArgAsnLysArgTyrGluLeuArgAlaGluTyrLysGlnProPheProGlyPhe 348
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 GTTTCGGTCCACAGCTTTTATACAAAGAGACACTCCAGAGCCCTGTATTAGTGCC 309
Qy 349 GluAlaLeuArgValHisLeuAsnArgAsnAspTyrHisHisAspGluLysAlaGlyAsp 368
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
310 AAGGCT-----ATCCAGCTGAACAGCAATAGTGTCAAGCCTTACTCTTAAAGGGGCA 363
Qy 369 AlaValGluAsnPhePheAsnGlnThrGlnAsnAlaArgIleGluLeuArgHis--- 387
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
364 GCATTAAAGAACATG-----GGAAGGTCGAGGAAGCATAATTCATTCCGAGAAGCA 417
Qy 388 -----GlnProIleGlyArgLeuLysGlySerTrpGlyVal-----GlnTyrLeu 402
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
418 ATTCGATTAGCACCTTGT---CGCTTGACTGCTGTGAAGTCTCATTGAATGCTATTTG 474
Qy 403 GlyGlnLysSerSerAlaLeuSerAlaThrSerGluAlaValLysGlnProMetLeuLeu 422
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 -----GCATCCAGCAGTGTGCGGAGCCATGCTGTGATG 507
Qy 423 AspAsnLysValGlnHisTyrSerPhePheGlyValGluGlnAlaAsnTrpAspAsnPhe 442
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
508 GCCAACAAATGTC-----TACAAACCCCTTGA-----GCCAACGCCAACAACTCTA 552
Qy 443 ThrLeuGluGlyGlyValArgValGlu-----LysGlnLysAlaSerIleArg 458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
553 ACTCTGTGGCACTGTGCTGGAATCCTGTTACCCAGAGAAAGCAAAACATTG 612
Qy 459 TyrAspLysAlaLeuIleAspArgGluAsnTyrTyrLysGlnProLeuProAspLeuGly 478
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
613 TTAGACAAAGCGCTCAITTCAGAGACCGGATTATATAAAGGCTGTGTGAA-GAAAGCGGA 671
Qy 479 AlaHisArg----- 481
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
672 GTTACTAAGTCGAGAGAGAGAGTATGAAGAGGAATTGCCCTGCTGCGTATACCTTGC 731
Qy 482 -----GlnThrAlaArgSerPheAla----- 488
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
732 AAACCAAGTACTGCGTACTACTACATCGATGCTAGGAGATTTTCTGGTGGCAGTTAATGA 791
Qy 489 LeuSerGlyAsnTrpTyrPheThrProGlnHisLys-LeuSerLeuThrAlaSerHisG1 508
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
792 GTTTCAGGAAGC---TATGGATCAGTACAGCATCGCCTCAGTTTGGATCCAAATGACCA 848
Qy 508 nGluArgLeuProSerThrGlnGluLeu 517
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
849 AAAATCCTTGAGGGAATGCAGAAAATG 876
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 11:00:13 ; Search time 36 Seconds

(without alignments)
3843.891 Million cell updates/sec

Title: US-09-936-377-2

Perfect score: 4036

Sequence: 1 MAQTLKPIVLSILLINTPL.....FLSDTPQMGRSFTGGVNVKF 758

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3845	95.3	764	15	US-10-066-551-2
2	216	5.4	713	11	US-09-801-451A-4
3	214.5	5.3	720	15	US-10-066-551-3
4	202	5.0	713	11	US-09-801-451A-2
5	198.5	4.9	687	12	US-10-181-319-2
6	185	4.6	660	12	US-10-238-075-732
7	181	4.5	654	12	US-10-238-075-527
8	180.5	4.5	914	15	US-10-043-344-11
9	177.5	4.4	713	12	US-10-238-075-263
10	163	4.0	913	15	US-10-043-344-5
11	161.5	4.0	912	15	US-10-043-344-7
12	161.5	4.0	912	15	US-10-043-344-9
13	160	4.0	540	12	US-10-043-344-2
14	156	3.9	971	15	US-10-098-808-2
15	155.5	3.9	682	12	US-10-238-075-1077

16	155	3.8	908	15	US-10-043-344-94	Sequence 94, Appl
17	154	3.8	813	9	US-09-815-242-5085	Sequence 5085, Ap
18	154	3.8	911	15	US-10-043-344-107	Sequence 107, App
19	151.5	3.8	915	12	US-10-240-218-6	Sequence 6, Appli
20	150.5	3.7	915	9	US-09-332-226-2	Sequence 2, Appli
21	150.5	3.7	915	15	US-10-043-344-96	Sequence 96, Appl
22	147.5	3.7	909	9	US-09-332-226-4	Sequence 4, Appli
23	147.5	3.6	911	15	US-10-043-344-95	Sequence 95, Appl
24	141.5	3.5	729	9	US-09-815-242-10132	Sequence 10132, A
25	135	3.3	791	10	US-09-881-752A-334	Sequence 334, App
26	127	3.1	815	9	US-09-815-242-5106	Sequence 5106, Ap
27	125	3.1	793	12	US-09-881-752A-362	Sequence 362, App
28	123	3.1	793	12	US-09-882-227-78	Sequence 78, Appl
29	122.5	3.0	767	10	US-10-180-326-1	Sequence 1, Appli
30	122.5	3.0	1590	15	US-10-180-326-1	Sequence 1277, Ap
31	118.5	2.9	753	12	US-10-238-075-1277	Sequence 2, Appli
32	117.5	2.9	1475	9	US-09-740-274-2	Sequence 5247, Ap
33	117.5	2.9	422	9	US-09-815-242-5247	Sequence 5456, Ap
34	117	2.9	978	9	US-09-815-242-5456	Sequence 12686, A
35	115.5	2.9	1001	9	US-09-815-242-12686	Sequence 12211, A
36	112	2.8	981	9	US-09-815-242-12211	Sequence 62, Appl
37	111	2.8	631	15	US-10-260-877-62	Sequence 6493, Ap
38	109.5	2.7	818	14	US-10-055-364-44	Sequence 44, Appl
39	109	2.7	121	14	US-10-022-461-5	Sequence 5, Appli
40	108	2.7	907	12	US-10-130-973A-5	Sequence 5943, Ap
41	107.5	2.7	430	15	US-10-156-761-7943	Sequence 7943, Ap
42	106	2.6	511	15	US-10-122-706-30	Sequence 30, Appl
43	106	2.6	970	15	US-10-331-061-5	Sequence 5, Appli
44	106	2.6	1052	12	US-10-130-973A-17	Sequence 17, Appl
45	106	2.6	1172	15	US-10-122-706-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-10-066-551-2
; Sequence 2, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Edwards, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the prevention and treatment of Neisserial infections
; FILE REFERENCE: 875.045US1
; CURRENT APPLICATION NUMBER: US/10/066,551
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-066-551-2

Query Match 95.3%; Score 3845; DB 15; Length 764;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 730; Conservative 7; Mismatches 18; Indels 6; Gaps 3;
OY 1 MAQTLKPIVLSILLINTPLLAQAHETEQSVGLTWTWVKSRPRATSGLLHTSTASDKI 60
DB 7 MAQTLKPIVLSILLINTPLLSQAHTGEQSVGLTWTWVKSRPRATSGLLHTSTASDKI 66
OY 61 ISGDTLRQKAVNLGDALDGVPGIHASQYGGASAPVIRGTGRIRKVLNHHGTGMADF 120

Db	67	ISGDTLRKAVNLGDALDGVPEIIHASQVGGASAPVIRGQTRRIKVIKLNHHGSETGDMADF	126
QY	121	SPDHAIWDTALTASQOVEILRGPGVTLLYSNGNVAGLVDAWGKIPKXMPENGVSGBELGURL	180
Db	127	SPDHAIMVDSALSQOVEILRGPGVTLLYSNGNVAGLVDAWGKIPKXMPENGVSGBELGURL	186
QY	181	SSGNLEKLTSGGINTGLGKNFVLHTEGLYKSGDYAVPRYRNLIKRLPDSPPRFANGQHRA	240
Db	187	SSGNLEKLTSGGINTGLGKNFVLHTEGLYKSGDYAVPRYRNLIKRLPDS--HAUSQTS	243
QY	241	V-LGW--RKRFRYRYSRRDOYGLPAHSHEYDDCHADIITWQSLINKRYLOLYPHLLTE	297
Db	244	IGLSWVGEGKFTGAAYSRRDOYGLPAHSHEYDDCHADIITWQSLINKRYLOLYPHLLTE	303
QY	298	EDWDYDNPGLSCGFEHDDDDAHAAHNGKPMIDLENKEYELRAEWKOPFPFGFALRVLHNR	357
Db	304	EDIDYDNPGLSCGFEHDDDDAHAAHNGKPMIDLENKEYELRAEWKOPFPFGFALRVLHNR	363
QY	358	NDYHDEKAGDAVENFENNONTONARIELRHOPICRLKSGMGVOYLQOKSSALSATSEAVK	417
Db	364	NDYRDEKAGDAVENFNNONTONARIELRHOPICRLKSGMGVOYLQOKSSALSATSEAVK	423
QY	418	QPMLLDNKVQHYSPFVGBOQANDNFTLBGGVVRVKQKASIRYDKALIDRENYKQPLPDL	477
Db	424	QPMLLDNKVQHYSPFVGBOQANDNFTLBGGVVRVKQKASIRYDKALIDRENYNHPLPDL	483
QY	478	GAHQQTARSPALSGNWYFTPOHKLSLTASHOERLPSTQELYAHGKHVATNTPFVGKNKHLN	537
Db	484	GAHQQTARSPALSGNWYFTPOHKLSLTASHOERLPSTQELYAHGKHVATNTPFVGKNKHLN	543
QY	538	KERSNNIELALGYEGDRWOYNALYRNBFGNYIAQTLNDGRGPKSIDDESEMKLRYNQ	597
Db	544	KERSNNIELALGYEGDRWOYNALYRNBFGNYIAQTLNDGRGPKSIDDESEMKLRYNQ	603
QY	598	SGADFYGABGEIYFKPTPRYIRIGVSGDYVRGLKNLPSLPGREDAYGNRPFIAQDDQNAP	657
Db	604	SGADFYGABGEIYFKPTPRYIRIGVSGDYVRGLKNLPSLPGREDAYGNRPLIIAQADQNAP	663
QY	658	RVPAAELGHLKASLTDIRIDANLDYRYFAQNKARVETRTPGHMLNLGANYRNRTRYG	717
Db	664	RVPAAELGHLKASLTDIRIDANLDYRYFAQNKARVETRTPGHMLNLGANYRNRTRYG	723
QY	718	ENWYVKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF	758
Db	724	ENWYVKADNLLNQSVYAHSSFLSDTPQMGSRFTGGVNVKF	764

RESULT 2

US-09/801-451A-4
Sequence 4, Application US/09801451A
Publication No. US20030096368A1
GENERAL INFORMANT:
APPLICANT: Sparling, P. Frederick
Beucher, Margaret
TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule
and Vaccine
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: InClone Systems Incorporated
STREET: 180 Varlick Street
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/801,451A
FILING DATE: 08-Mar-2001
CLASSIFICATION: <Unknown>

Db 617 NPNLEIGHGRVY-----QKATGSLAAGQKDRGKLENVVRQGFVN----- 659
 QY 674 DRIDANLDYRVPQAKNKLARYETRTPGHMLMLGANYRNTRYGWNWTVKADNLLNSQV 733
 Db 660 -----DVPANMKPLGKOTLN-----VNLNVN-----NVDFKPY 687
 QY 734 YAHSSFLSDT-POMGRSFTGGVNVKF 758
 Db 688 YPHSQRTWNTLPGVGRDVRGLGVNYKF 713

RESULT 3

US-10-066-551-3
 ; Sequence 3, Application US/10066551
 ; Publication No. US20030100071A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Apicella, M. A.
 ; APPLICANT: Edwards, J. L.
 ; APPLICANT: Gibson, B. W.
 ; APPLICANT: Scheffler, K.
 ; APPLICANT: Brown, E.
 ; TITLE OF INVENTION: Vaccine and compositions for the
 ; prevention and treatment of Neisserial infections
 ; FILE REFERENCE: 875.045051
 ; CURRENT FILING DATE: 2002-01-31
 ; PRIOR APPLICATION NUMBER: US 60/344,452
 ; PRIOR FILING DATE: 2001-10-23
 ; PRIOR APPLICATION NUMBER: US 60/310,356
 ; PRIOR FILING DATE: 2001-08-06
 ; PRIOR APPLICATION NUMBER: US 60/266,070
 ; PRIOR FILING DATE: 2001-01-31
 ; NUMBER OF SEQ ID NOS: 12
 ; SEQ ID NO 3
 ; LENGTH: 720
 ; TYPE: PRT
 ; ORGANISM: Neisseria gonorrhoeae
 US-10-066-551-3

Query Match 5.3%; Score 214.5; DB 15; Length 720;
 Best Local Similarity 21.6%; Pred. No. 3.6e-11;
 Matches 186; Conservative 101; Mismatches 315; Indels 259; Gaps 45;

QY 16 INTPL-----LQAQTEQSGVLETVTVVVKSRPRATSGLLHTSTASDKIIS 62
 Db 1 MNTPLFRLSLSLAAGFAAENNAKVLDITVIVKGRQ-----GSKRTIV- 50
 QY 63 GDTLRQK-----AVNLGDALDGVPGIHASQYGGGASAPVI-----RGQTGRRIKY----- 107
 Db 51 --TLQXDESTATDMRELLKEPSI---DFGGNGTSGQFLTLRGWGQNSVDIKVDNAYS 105
 QY 108 ---LNHGETGDMADFPDHAIMVDYDALSQOVELRGVTVLLYSSGNVAGLVADVADGKIP 164
 Db 106 SQILYHQGR-----FVDPALVKKVVSQKQ-----AGSASAGIGATNGAII 146
 QY 165 EKMPF-----NGVSGELGLRSSG--NLEKLTSGGINIGLKNFVLTEGLY-----RKSG 213
 Db 147 AKTVDAQDLLKGLDKNMGVRLNSGFASNEGVSYGASVFGKEGNF---DGLFSYNRNDEK 202
 QY 214 DYAVPR-VRNIL---KRLPDSP-----RRFANGQRAVLGWRKRYFR----- 251
 Db 203 DYEAGKGFNVNGKGTVEYSALDKRSYLAKIGITTFGDDHRLVLSHMKDQHRGIRTVREE 262
 QY 252 -TYSRRDQYGLPAHSHEYDDCHADIIWOKSLINKRVLYLPHLLTEEDVDYDNFGLSCG 310
 Db 263 FTVGKSSRINDROAPAYR-----TTQSNTLWATGKNLG 299
 QY 311 FHDDDDAHAAHNGKPFIDLNKRY---ELRAEWKQPPGPEALRVHLNRNDYHHDKAG 367
 Db 300 FVEKLDANAYV-----LEKERYSDDSGTGAGNVKGFNHTTRTTRGANFNFDGRLA 351
 QY 368 DAVENFFNQT-QNARIELHQ---PIGLKSGMGVQVYLGOKSSALSATSEAVKQPM--- 420

Db 352 E-----OTLLKYGINVRHQEIKPOAFLNSKFSIPTBEKN-----GQKVDKPMEOQ 397
 QY 421 -----LNDKVKQHYSPFGVEQA-NMDNFTLEGSVRVEKOKASIRYDKALIDR 466
 Db 398 MKRADEDTWHAYKLSNPTTIDTGVYVEAHDGFTLTGGLRYDRFKVK-THDGKTVSS 456
 QY 467 ENYTKQPLDLAGHARTARSFALSNGWYFTPOHKLSTASHQERLPSTOELYAHGKH--- 523
 Db 457 SNLN-----PSFGVIWQPH-----HWSFSASHNY---ASRSPRL---YDALQTHCKRGII 501
 QY 524 -VATNTFEVGNKHLNKNERNNIELALGYEGDRWQYNLALYENRGNIIYIYATLNDGRPK 582
 Db 502 SIADGT-----KAERARTEIGFNYNDGTFAN-----GSYFW-OTIKDALANP 544
 QY 583 STEDDSEMKLVRYNQSADFYGAE-GEIYFKPTPRYRIGVSGDYVRGLKNLPLSPGRED 641
 Db 545 QNRHDSVAVREAVNAGYIKNHGYELGASRTGGLTAKGVVS-----HSKPRFYD 593
 QY 642 AYGNRPFTIAODDQNAAPRVPAARLGHFLKASITDRI-DANLDY-YRVFAONKLARYETRT 699
 Db 594 THKDKLLSANPEF-----GAQVGRTWTSALAYRFQNPNIIGWR-----GRYVOKAT 640
 QY 700 GHMMLNLCANVR-----NTRYGEM-----NWYVKADNLLNQSVYAHSS 738
 Db 641 G-SILAAQCKDKRKNLENVVRKGFVNDVPANVPLGKDTLNVLSVNNVFNKPYPHSQ 699
 QY 739 FLSDT-POMGRSFTGGVNVKF 758
 Db 700 RWTNTPGVGRDVRGLGVNYKF 720

RESULT 4

US-09-801-451A-2
 ; Sequence 2, Application US/09801451A
 ; Publication No. US20030096368A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sparling, P. Frederick
 ; TITLE OF INVENTION: Isolated PrpB Nucleic Acid Molecule
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ImClone Systems Incorporated
 ; STREET: 180 Varick Street
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10014
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION NUMBER: US/09/801,451A
 ; FILING DATE: 08-Mar-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/418,964
 ; FILING DATE: 07-APR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gallagher, Thomas C.
 ; REGISTRATION NUMBER: 37,066
 ; REFERENCE/DOCKET NUMBER: SPA-3-P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 645-1405
 ; TELEFAX: (212) 645-2054
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 713 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

326 SHTLIPGHAQUSYGAIVYRQOQKPEGSAT-----LYPEGNI DFTSLYFQDQMTKSY 379
444 LBGVVRVEXKQASIRYKALIDRENYKOPLP---DLGHRQTAARSFALSGNWTPOHK 500
380 VNIIV-----GSRYDR-----YKGFNPRAGELKAERLSPRA-AIS-----VSPDTW 419
501 LSL--TASQERLPSQELVYAHGKHVATWTFEYGN-----KHLNERSNNIELALG--- 549
420 LMVYSISSAFRAPTWAEYRDDVHP-----YKGFNYPVWNLNPKENNITREIGAGIQ 475
550 -----YEGDRWQYNLALYRNFRGNYIYAQTLDNDRGPKSIEDDSEMKLVRYNOSGADPYG 604
476 LDGLLTNDRLQLKGVFGTDARNYIATRV-----DMKMRSSYSYNVSRARITWG 524

RESULT 8
US-10-043-344-11
; Sequence 11, Application US/10043344
; Publication No. US2003008086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Mordin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 11
; LENGTH: 914
; TYPE: PRP
; ORGANISM: Haemophilus influenzae
US-10-043-344-11

Query Match 4.5%; Score 180.5; DB 15; Length 914;
Best Local Similarity 19.7%; Pred. No. 9.8e-08;
Matches 200; Conservative 132; Mismatches 322; Indels 359; Gaps 52;

QY 5 TLKPI-----VLSILLINTPLLAQAH-----ETEQSVG-----LETVTVVG-KSRP 44
Db 2 TKKPYFRLSIISCLISYVKAETQSIKDTKEAISSEVDQSTEDSELETISVTAEKIRD 61
QY 45 RATSGLLHTTASDKII--SGDITLROKAVNLGDVPGIHASOYGGGASAPV-IRGQT 101
Db 62 KXNEV-----TGLKIIKTSISISREQVILNIRDLTRYDPGISVVEQGRGASSYSIRGMD 117
QY 102 GRIKVLNHHGTGDMADFPDHAIMVDYDTSALSOQVEILRGPVTL--YSGSNVAGLVDA 159
Db 118 RNRV-----ALLVDGLPQTSYVQVQSVPLVARSYSGTGAINIEYIE 158
QY 160 DGKIPKMPENGVSSELGLRLSSNLEKLTSGGINIGLG-KNFVLHTEGLYR----- 210
Db 159 NVKAVE-ISKGSSEYNGALAGSVTFQSKSAADILEGDKSWGIQTKNAYSNNKGFTH 217
QY 211 -----KSGDY---AVPRYNLKLPSDPRFANGQHR-AVLGWRKFFRYRTYSDRRDQY 260
Db 218 SLAVAGKGGFDGVAITYRN-----SIEQVHKDALKGVS--YHRLIAKPEDQ- 265
QY 261 GLPAHSHEYDDCHADITWQSL-INRYLQIYPHLLTEEDVDY-----DNP----- 305
Db 266 --SAYFMQDECPKDDYNSCLPFAKRPAILSSQRTVSVSDYTGANRIKPNMKEYESQS 323
QY 306 -GLSCGFHDDDAHAHANGKFWDLNKRVELRAENKQPPGPFALRVHLNRYNDYHDE 364
Db 324 WFLRGYHVFSEQ---HYGG--IFEFTQKFDIR---DMTFPAY-----LRSTE 364

QY 365 KAGDAVENFENNOTQA--RIELRHOPIGR-----LKGSMGVOVLGOKSSA 408
Db 365 KRDDSSGTFPKQDYGAYQRIE-----DGRGVNYSGLIYDEHHRKORGVIEIYENKNK 419
QY 409 LSATSEAV---KQPMLLDNKVQH-----YSPFGVE----- 435
Db 420 AGIIDKAVLSANQONIIILDSYMQHTHCSLYPNPKNCRPTRDKPYSYHSDRNRYKCKHN 479
QY 436 -----QANW-----DNFT-----LEGGVRVEKQKASIRYDKALIDR 466
Db 480 MLQINLEKKIQONWLTHQIVFNLFDDFTSALQKHVLTTRVTATAKIS--EKANETR 536
QY 467 ENYKQ-----PLPDLG-----AHRQTARSPALSGNWTFTPOHKL 503
Db 537 RNYKKQPYLPKPTVGVVQDCHDYKGNSSNYRDCVKRLIKGNYYFAARNNNALGKYV 596
QY 504 -----TASQOE-----RLPSTQEL 517
Db 597 DLGLGIRYDVSRTKANESTISVGKFKNFSWNTGIVIKPTWMLDSYRLSTGFRNPSFAEM 656
QY 518 YAHGKHVATNTFEVGNKHLNERSNNIELALGYEGDRWQYNLALYRNFRGNYI-YAOTLN 576
Db 657 Y--GWRVGGNNSVYVGVKFKPETSBNQEPGLALKGDFGNIESHFSNAYRNLIATAFELN 714
QY 577 ---DGRPKSIEDDSEMKLVRYN--OSGADFYGABGEIYF----- 611
Db 715 KNGTGKANYGHNAQNAKLVGNITAQDLFNLGLWKRIPIYGYWATFAYNRVVKVDQKINAG 774
QY 612 -----KPTPRYIGVSGD-----YVRGLKNLPSLPGREDAYGNRPF 648
Db 775 LASVSVLFDALQFS--RYIIGLYDHSNTWGTINTMETQSKAKSONELLGKR-ALGNN-- 830
QY 649 IAQDDONAPRVPAARLGFHLKASITDRIDANLDYRVFAQNKARYETRETEGHMLNIGA 708
Db 831 -SRDVKSTRKLTTRA---WHI-----LDVSGYGMANK-----NIMRLGI 865
QY 709 NYRNRTRYGENWYVK-ADNLLN--QSVYAHSSFLSDTPQMGSRFTGNNVVF 758
Db 866 YNLFNRYRYTWEAVRQTAQGAVNQHVNGSYTRYAAS-----GRNYITLTLEMKF 914

RESULT 9
US-10-238-075-263
; Sequence 263, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol
; of E.coli, and biological uses of these polynucleotides and of th
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 263
; LENGTH: 713
; TYPE: PRP
; ORGANISM: Escherichia coli
US-10-238-075-263

Query Match 4.4%; Score 177.5; DB 12; Length 713;
Best Local Similarity 19.5%; Pred. No. 1.3e-07;
Matches 167; Conservative 103; Mismatches 283; Indels 305; Gaps 41;

QY 6 LKPIVLSILLINTPLLAQAHETEVSGL-----TWTVVGKSRPRATSGL--LHTSTA--- 56
Db 11 LKGRALFSLFAAPMI---HATDSVTTKOGETIIVTADANTATEATDGYQLSTATLT 67
QY 57 -----SDKIIISGDTLRQKAVNLGDALDGVPG-IHASOYGGGASAPVIRGQT 101
Db 68 DMPMLDIPQVNVTVSDQVLE-----NQNATTLDEALYNSVNVQNTLGGTODAFVRRGFG 123

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 913
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
; US-10-043-344-5

Query Match
Best Local Similarity 4.0%; Score 163; DB 15; Length 913;
Matches 170; Conservative 143; Mismatches 304; Indels 352; Gaps 48;

Qy 102 GRIKVLNHHGTGMDAFSPDHAIWV-----TALSQOQVEILRGVPTLLYSSGN 151
Db 124 ANR-----DGSMTWGLRTVLPSPNAATERVEVLKGPASTLYGLLD 165
Qy 152 VAGLVADVADGKIPIKMPENGVSGLGLSSGNLEKLTSGGINIGLKG---NFVLHTEG 207
Db 166 PGLINVT-KRPEKTFHGSVA-----TSSFGGTGQLDITGPIEGTQL 210
Qy 208 LYRKSGDVAVPYRNKRLKPLDSPRPANGQRAVLGWRKFRFYRYSDRRQYGLPAHSH 267
Db 211 AYRLTCEV-----QDEDYWRNFGKERSTFIAPS--- 238
Qy 268 EYDDCHADIIWQSLINKRYLQLYPH-----LITBEDVD-----YDNFGL 307
Db 239 -----LTFVGD-NATVTMLYSHRDYKTFPDRGTIFDLTKQPVNVDRKIREDEP- 286
Qy 308 SCGFHDDDAHAHANGKFWIDLRNRYELRAEWQPPGPEALRVHLNRNDYHHDEKAG 367
Db 287 ---FNITD-----GQSDLAQLNAEYHLNSQWTARF-----DYSYSQ--- 319
Qy 368 DAVENFFNQ-----TONARIEL-----RHQPIGRKLGSGVQVYLGOKSSAL 409
Db 320 ---DKYSDQARVAYDAITGTLTRVVDATQSTQRMHATRADLOGN-VDIAGFYNEIL 374
Qy 410 SATS-----EAVQPMILLNNKQVHYSPFG-----VEQANW----- 439
Db 375 GGVSYEYDILRTDMIRCKKAKDFNIYNPVYNTSKCTVTSASDSQOTIKOENYSAQAQ 434
Qy 440 -----DNFTLEGVRVEKQKASIRYDKAL-IDRENYKQPLDPLGAHRTARSPALSGNW 493
Db 435 ALVLTNDWIAVAGRIQYTYTAGKGRPFNVNVDSDRDEQWTPALGLVYKLTSPSVLFANY 494
Qy 494 --YFTPOHKLSTASHQERLPSTQELYAHGKHVATNTFEVGNKHLNKRNNIELALGYE 551
Db 495 SQTFFMQ---SSIASYIGDLPE-----SSNAYEVGAK-----FELFDGIT 532
Qy 552 GDRWQYNLALYRNRFNGNLYAQTLDN-----GR-GPKSIEDSEMKLVRYNQSGADPY 603
Db 533 AD-----IALFDIHKRNVLVTESIGDETAKTAGRVSRGVEVDLAGALTEMINIIASVG 587
Qy 604 GAEGEYFFKTPRYRIGVSGYVRGLKNLPLSPGREDAYGNRPFTAQDDQNAAPRPAAR 663
Db 588 YTDKAVLEDP-----DYAGKPLNVPRHGT-----SLFLTVDIHNMPGNNTIT 630
Qy 664 L--GFH--LKASLTDRIDANLDYRYVFAQNKLARYETRTPGHHMLNLGANYRNTRYGEW 719
Db 631 FGGGHHGVSSRSATN-----GADYY-----LPGYFVADAFAPAAAYKMKLQY-PV 671
Qy 720 NWYVKADNLINQSVYVAHS 737
Db 672 TLQLNVKNLFDKTYTTS 689

RESULT 10
US-10-043-344-5
; Sequence 5, Application US/10043344
; Publication No. US2003008086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Mordin, Andrew D.
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
```

Db 779 SYLFDALQPSRYIIGLYDHPSTNT-----WGIK-----TMFTQSKAKSONELLGKRA 825

Qy 751 TG--GVNVK 757

Db 826 LGNNSRNVK 834

RESULT 11

US-10-043-344-7

Sequence 7, Application US/10043344

Publication No. US2003008086A1

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Harkness, Robin E.

APPLICANT: Schryvers, Anthony B.

APPLICANT: Chong, Pele

APPLICANT: Gray-Owen, Scott

APPLICANT: Murdin, Andrew D.

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES

FILE REFERENCE: 1038-1221 MIS

CURRENT APPLICATION NUMBER: US/10/043,344

CURRENT FILING DATE: 2002-07-01

PRIOR APPLICATION NUMBER: 08/649,518

PRIOR FILING DATE: 1996-05-17

NUMBER OF SEQ ID NOS: 160

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 7

LENGTH: 912

TYPE: PR

ORGANISM: Haemophilus influenzae

US-10-043-344-7

Query Match 4.0%; Score 161.5; DB 15; Length 912;

Best Local Similarity 19.6%; Pred. No. 6.5e-06;

Matches 200; Conservative 129; Mismatches 315; Indels 377; Gaps 54;

Qy 5 TLKPI-----VLSILLINTPLAQAH---ETQSVG-----LETVTVG-KSRP 44

Db 2 TKKPYFRLSIISCLLSICVYKAEQTSIKDKTEAISSEVDQTSTEDSELETISVTAEKIRD 61

Qy 45 RATSGILLHTSTASKII--SGDTRQKAVNLGDALDVGPIGTHASQYGGGASAPV-IRGQT 101

Db 62 RKDNV-----TGLKLIKTSISREBQVNLRLTRYDPGLSVVEQGRGASSGVSIRGMD 117

Qy 102 GRRIKV-----NHHGTGDMADSPDPHAIMVDLTALSOQVEILRGP 142

Db 118 RNRVALLVDGLPQTQSYVVSPLVARSGYSGTGAINIEYENV-----KAVEISKGG 169

Qy 143 VTLLYSSGNVAGLV-----DVADG-----KIPKMPENGVSCELGRLSSGNLE 186

Db 170 SSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTKNAYSNNKNGFTHSLAVAGKQGGFE 229

Qy 187 KLTSGGINIGLKNFV---LHTEGL-----YRKSQDY----- 215

Db 230 -----GLAIYQRTSIEQVHKDALKGVQSDYRLIATTDKSSGYFVIGQECNPGDDKCA 284

Qy 216 -----AVRYRNKRLPDSPPRAN-----GQH---RAVLGWRKRFYRRT 252

Db 285 KPATLTQSTVSDVTGANIKPNEMKYESQWFLRGYHFSHQYIGGIFEFTQOK 344

Qy 253 Y-----SDRRD-----QYGLPAHSHE-----YDCH----- 273

Db 345 FDIRMTFPAYLSPATERDDSRFPYMQDHGAYQHIEDGRGVKYASGLYFDEHHRQVR 404

Qy 274 -----ADII-----WOKSLINKYLO-----LYPHLLTEEDVDYDNPGLSCG 310

Db 405 GIEYIYENKAKIGIDKAVLSANQQNIILDSYMRTHCSLP-----NPSKNCR 453

Qy 311 -----FHDDDDAHANHGKFWIDLRNKKRYELRAEW---KQPPP-GFEALRVHLNR 357

Db 454 PTLDKPYSYRSRDNVYKEKN---MLQI-NLEKKIQQWLTHQIVNFGLEDFTSALQH 509

Qy 358 NDYHHEKAGDAVENFFNNQTONARIELRHQPIGRKLGSGWGVQYLGOKSALSATSBAVK 417

Db 510 KDY-----LTFRVATATADSIPRKPGETGKPRNGLOS----- 540

Qy 418 QPMLLDNKKVQHYSPFQVEQANWDFLLEGGVVRVEKQKASIRYDKA---LIDRENY--- 470

Db 541 QPYLYP-KPEPY-FAGQDHCNY-----QSSSNRYRDKVRLIKGNYYFAAR 585

Qy 471 ---KQPLPDLG-----AHROTA-----RSFALSGNMYFTPOHKLISLT--ASHQ 508

Db 586 NMALGKYVDLGLGIRVDVSRTKANESTIISVGKFNFSWNTGIVIKETEWLDLSYRLSTG 645

Qy 509 ERLPSTQELYAHGXHVATNTFEVGNKHLNKNERSNNIELALGYEGDRQYQYMLALYRNFQ 568

Db 646 FRNPSFSEMY--GWRYGGKNDVVYGVFKPETSRSQEFGLAKGDFGNIEISHFSNAYRN 703

Qy 569 YI-YAQTIN---DGRGPKSIEDDSEMKLVRYN-QSGADFYG-----AEGEY 610

Db 704 LIAFABELSKNGTKGNVGYHNAQNAKLGVNITAQDLDFNGLWKRIPIGYWATFAYNQVK 763

Qy 611 FK-----PTPRYRIGVSGD-----YVRGRKLNPLSPFGRE 640

Db 764 VKDQKINAGLASVSYLFDALQPSRYIIGLYDHPSTNTGINTWFTQSKAKSONELLGKR 823

Qy 641 DAYGNRPFIADDDQNAQPRVPAARLGFHLKASLTDRIIDANLDYRVFAQNKLYRTEYTPG 700

Db 824 -ALGNN---SRDVKSTRKLTRA---WHI-----LDVSGYYWANK----- 855

Qy 701 HHMLNLGANYRNTRYGEMWVVK-ADNLLN--OSVVAHSSFLSDTPQMGRSFTGVNVK 757

Db 856 NIMRLGIYNLFNYRYVTWEAVRQTAQAVNQHQNVGSYTRYAAS---GRNYTLTLEMK 911

Qy 758 F 758

Db 912 F 912

RESULT 12

US-10-043-344-9

Sequence 9, Application US/10043344

Publication No. US2003008086A1

GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Harkness, Robin E.

APPLICANT: Schryvers, Anthony B.

APPLICANT: Chong, Pele

APPLICANT: Gray-Owen, Scott

APPLICANT: Murdin, Andrew D.

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES

FILE REFERENCE: 1038-1221 MIS

CURRENT APPLICATION NUMBER: US/10/043,344

CURRENT FILING DATE: 2002-07-01

PRIOR APPLICATION NUMBER: 08/649,518

PRIOR FILING DATE: 1996-05-17

NUMBER OF SEQ ID NOS: 160

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 9

LENGTH: 912

TYPE: PR

ORGANISM: Haemophilus influenzae

US-10-043-344-9

Query Match 4.0%; Score 161.5; DB 15; Length 912;

Best Local Similarity 19.6%; Pred. No. 6.5e-06;

Matches 200; Conservative 129; Mismatches 315; Indels 377; Gaps 54;

Qy 5 TLKPI-----VLSILLINTPLAQAH---ETQSVG-----LETVTVG-KSRP 44

Db 2 TKKPYFRLSIISCLLSICVYKAEQTSIKDKTEAISSEVDQTSTEDSELETISVTAEKIRD 61

Qy 45 RATSGILLHTSTASKII--SGDTRQKAVNLGDALDVGPIGTHASQYGGGASAPV-IRGQT 101

Db 62 RKDNV-----TGLKLIKTSISREBQVNLRLTRYDPGLSVVEQGRGASSGVSIRGMD 117

Qy 102 GRRIKV-----NHHGTGDMADSPDPHAIMVDLTALSOQVEILRGP 142

Db 118 RNRVALLVDGLPQTQSYVVSPLVARSGYSGTGAINIEYENV-----KAVEISKGG 169

Qy 143 VTLLYSSGNVAGLV-----DVADG-----KIPKMPENGVSCELGRLSSGNLE 186

Db 170 SSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTKNAYSNNKNGFTHSLAVAGKQGGFE 229

Qy 187 KLTSGGINIGLKNFV---LHTEGL-----YRKSQDY----- 215

Db 230 -----GLAIYQRTSIEQVHKDALKGVQSDYRLIATTDKSSGYFVIGQECNPGDDKCA 284

Qy 216 -----AVRYRNKRLPDSPPRAN-----GQH---RAVLGWRKRFYRRT 252

Db 285 KPATLTQSTVSDVTGANIKPNEMKYESQWFLRGYHFSHQYIGGIFEFTQOK 344

Qy 253 Y-----SDRRD-----QYGLPAHSHE-----YDCH----- 273

Db 345 FDIRMTFPAYLSPATERDDSRFPYMQDHGAYQHIEDGRGVKYASGLYFDEHHRQVR 404

Qy 274 -----ADII-----WOKSLINKYLO-----LYPHLLTEEDVDYDNPGLSCG 310

Db 405 GIEYIYENKAKIGIDKAVLSANQQNIILDSYMRTHCSLP-----NPSKNCR 453

Qy 311 -----FHDDDDAHANHGKFWIDLRNKKRYELRAEW---KQPPP-GFEALRVHLNR 357

Db 454 PTLDKPYSYRSRDNVYKEKN---MLQI-NLEKKIQQWLTHQIVNFGLEDFTSALQH 509

```

Db 62 RKDNEV---TGLGKIITKTESIREQVNLNIRLTRYDPGISVWEQGRGASSGYSIRGMD 117
QY 102 GRRIKVL-----NHGETGDMADSPDHAIWVDLALSOQVILRGP 142
Db 118 RNRVALLVDGLPQTOSYVVQSVFLVARSYGSGTGAINIEYENV-----KAVISKGG 169
QY 143 VTLLYSSGNVAGLV-----DVAAG-----KIPEKMPENGVSGBELGLRLSGNLE 186
Db 170 SSSEYNGALAGSVTFQSKSAADILEGDKSWGIGTKNAYSSKNKGFTTHSLAVAGQGGFE 229
QY 187 KUTSGGINLGNFV---LHTEGL-----YKSGDY----- 215
Db 230 -----GLAIYTORNSIETQVHKDALKGVQSYDRLIATTDKSSGYFVIOGCECPNGDKCAA 284
QY 216 -----AVPRYRLKRLPDSRRFAN-----GOH-----RAVLGWRKRYFRRT 252
Db 285 KPATLSTOSETVSVDYTGANKIKENPMKYESQSWFLGGYHFSHQHVGIFGIFETQOK 344
QY 253 Y-----SDRD-----OYGLPAHSHE-----YDCH----- 273
Db 345 FDIRDMTFPAYLSPTEERRDSSRSFYPMQDHGAYQHIEDGRGVKYASGLYFDEHHRKQV 404
QY 274 -----ADII-----WOKSLINKRYLO-----LYPHLLTEEDVDYDNLGLSCG 310
Db 405 GIEYIYENKAGIIDKAVLSAQNIILDSYWRHTCSLYP-----NFSKNCR 453
QY 311 -----FHDDDDAHAAHNGKPIDLNRKRYELRAEW-----KQPPP-GFEALRVHLNR 357
Db 454 PTLDPKYSYVSRDRNVYKEKHN---MLQL-NLEKKIQONLTHQIVFNLFGLDFTSALQH 509
QY 358 NDYHHEKAGDAVENFNQNTQNAIELRHQPIGLKSGWGVQYLGQKKSALSATSEAVK 417
Db 510 KDY-----LTRVIATADSIPRKEGETGKPRNGLOS----- 540
QY 418 QPMLLDNKVQHSYFFGVEQANWMDNFTLEGVRVEKQKASIRYDKA---LIDRENY--- 470
Db 541 QPVLVP-KPEPY-FAGODHCNY-----QSSSNYRDCKVRLIKKNYFYAR 585
QY 471 -----KOPLPDLG-----AHRQTA-----RSFALSGNMYFPPQHKLSLT--ASHQ 508
Db 586 NNMALGKYVDLGLIGIRYDSRTANESTISVGKFKNFSWNTGIVIKPTEWLDLSVRLSTG 645
QY 509 ERLPSTCELYAHKHKVATNTFEVGNKHLNKNERSNNIELALGVEGDRWQVNLALYRNFEN 568
Db 646 FRNPSSEMY---GWRVGGKNDVYVKGFKPETSROQFGLALKDGCNIEIHSFNAYEN 703
QY 569 YI-YAOTLN---DGRGPKSIEDDSEMKLVRYN--QSGADFYG-----AEGEY 610
Db 704 LIAFABELSKNGTKGNYGYHNAQNAKLVGNITAOQDFNGLWKRIPIYGWYATFAYNQVK 763
QY 611 FK-----PTPRYRIGVSGD-----YVRGRKKNLPSLPGR 640
Db 764 VKDQKINAGLASVSVLFDIAQPSRYIIGLYDHPNTWGINMTFTQSKAKSONELLGKR 823
QY 641 DAYGNRPFIADQDQNAFRVPAARLGFHLKASLTDRIDANLDYRVPFAQNKARLYETRTPG 700
Db 824 -ALGN--SRDVKSTKLTTRA---WHI-----LDVSGYYWANK----- 855
QY 701 HMLNLICANYRNRTRGEEWNYVK-ADNLLN--QSVYAHSSFLSDTPQMGRSFTGGVNVK 757
Db 856 NIMRLIGIYLNLFNRYRYTWEAVRQTAQGVNQHNVGSYTRYAAS-----GRNYTLLEMK 911
QY 758 F 758
Db 912 F 912

```

RESULT 13

```

US-10-238-075-721
; Sequence 721, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.

```

RESULT 14

```

US-10-098-808-2
; Sequence 2, Application US/10098808
; Publication No. US20030007981A1
; GENERAL INFORMATION:
; APPLICANT: Potter, Andrew A.
; APPLICANT: Rioux, Clement
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
; TITLE OF INVENTION: TRANSFERRIN-BINDING PROTEINS
; FILE REFERENCE: 9000-0049.20
; CURRENT APPLICATION NUMBER: US/10/098.808
; CURRENT FILING DATE: 2002-03-13

```

```

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of th
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238.075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 721
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-721

```

Query Match 4.0%; Score 160; DB 12; Length 540;

Best Local Similarity 22.1%; Pred. No. 3.9e-06; Matches 122; Conservative 78; Mismatches 229; Indels 124; Gaps 31;

```

QY 209 YRKSQDYAVPRYRLMK--RLPDSRRRFANGQRAVLGWRKFRYRTYSDDRDQYGLPAHS 266
Db 27 YRINGSYSDQGNRDTDPGRLEPNT--NYRNNSSQGVWLVGYNNGNR--FGLSLDRYRLATQT 82
QY 267 HEYDDCHADITWOKSLINKRYLQLYPHLLTEE-----DVDYDNFGLSCGFHDDDAHAHA 321
Db 83 Y-YEDPDGSY-----EAFSVKIPKLEKREKVGFYDTDVDGYLK-----KIHFDA 126
QY 322 HNGKPIWDLNRKRYELRAEWKQPPFPEALRVHLNRNDYHDEKAGDAV--ENPFNNQQTQ 379
Db 127 YEQTIQRFAN--EVKTTQVPSEMIQALTVH-NKTD--THDKQYTOAVTLQSHFSLPAN 181
QY 380 NARIELRHQPIGLKSGWGVQY---LGQKSSALSATSEAVKQPMLLDNKVQHSYFFGVE 435
Db 182 NELVT-----GAQYKQDRVSRQSGMGTSSKSLTG--PINKETRTSRYSYESE 225
QY 436 QA-----NW---DNFTLEGVVRVEKQKASIRYDKALIDRENNYKQPLDLAGHROTA 484
Db 226 QSTVSLFAQNDWRADFHWITWTVGVR--QYWLSSKLTGR--DGVSYTAGIISDTSLARESA 281
QY 485 --RSPALSGNMYFPPQHKLSLTASHQER--LPSTQELYAHGKHVATNTFEVGNKHLNKR 540
Db 282 SDHEMVTSTSLRYSGFDNLELRAAPAGQYVFTLSQLFMQTSAGGSVTY--GNPDLKAEH 339
QY 541 SNNIELALGYEGDRWQVNLALYRNFYNYIYAOTLNDGRGPKSIEDDSEMKLVRY--NQS 598
Db 340 SNNFELGARYNGNQWLIDSAVYSEAKDYI--ASLICDGSIVCGNNTSSRSYYDYDNI 398
QY 599 GADFYGARGEITYFKPTPRYRIGVSGDYVRGRL-----KNLPSLPGR----- 639
Db 399 RAKTWGLE-----ISAENYGVFSPYISGNLIRROYETSLTKTNTGEPAINRIGLKH 452
QY 640 -----BDAYGNRRPFIADQDQNA--PRVPA-ARLGFHLKASLTDRIDANLDYRVRV- 685
Db 453 TLVMQANIIISDVFTRAASSAKDDSGTETVNPVGWATLNFVAVNTEF-----GNEDQYRIN 507
QY 686 FAQNKLA--RYET 696
Db 508 LALNNLTDKRYRT 520

```



```

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/405,728
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/267,749
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Haemophilus somnus
; US-10-098-808-2

Query Match      3.9%; Score 156; DB 15; Length 971;
Best Local Similarity 18.8%; Pred. No. 2.4e-05;
Matches 195; Conservative 120; Mismatches 366; Indels 354; Gaps 50;

QY 6 LKPIVISILLINIPLL-----AQAHETQSGLVETVTVGKSRPRNTSGLLHTSTAD 58
D 6 LKPIVISILLINIPLL-----AQAHETQSGLVETVTVGKSRPRNTSGLLHTSTAD 58
QY 9 LKLIITLAVSTIFLPTTEAVADTESPSNTEAVLEIAIQVQAK-----HEISRHD 58
D 9 LKLIITLAVSTIFLPTTEAVADTESPSNTEAVLEIAIQVQAK-----HEISRHD 58
QY 59 KIISG-----DTLROKAVNLGDALDGVPGIHASQYGGGA-SAPVIRGQTRRIKVL 108
D 59 KIISG-----DTLROKAVNLGDALDGVPGIHASQYGGGA-SAPVIRGQTRRIKVL 108
QY 59 NEVTGLKVKVKSSEDIDKELLILNRLTRVDPGLSVVQGRGATSGYAMGVDRNRVAML 118
D 59 NEVTGLKVKVKSSEDIDKELLILNRLTRVDPGLSVVQGRGATSGYAMGVDRNRVAML 118
QY 109 -NHGETGMADFPD---HAI-MVDTALSOQVEILRGPVTLIYSSGNVAGLVADGKI 163
D 109 -NHGETGMADFPD---HAI-MVDTALSOQVEILRGPVTLIYSSGNVAGLVADGKI 163
D 119 VDLGQAQSYSTLKSANGGAINIEIYENIKSIELSKSSSAEYSGALGAGVGFRTKEA 178
QY 164 PEKMPEN---GVSGELGRLSSNLEKLTSGGINIG-----LGNFVLH----- 204
D 164 PEKMPEN---GVSGELGRLSSNLEKLTSGGINIG-----LGNFVLH----- 204
D 179 DDVIEQONGLNKATYKSNKSOFTQSVAGAFVGGFDSLAIFTRKKGKTRVHAABE 238
QY 205 -----TEGLYRKSGDYAVPRVNLK-----RLPSPRPFANGQHRV- 241
D 205 -----TEGLYRKSGDYAVPRVNLK-----RLPSPRPFANGQHRV- 241
D 239 IQHTYQPLEGVFNR---YEVDQNRNGKPVLANAYVILADECSNLSDPSCRHAKAKTRVG 295
QY 242 -----LCW-----RKGFYRTYS-----DRDQYGLPAHSEY----- 269
D 242 -----LCW-----RKGFYRTYS-----DRDQYGLPAHSEY----- 269
D 296 APENPNWTEEQAAAMPYPTKASAKDYTPDRISPNDMDYQSHSFFWKGYYRLSPN 355
QY 270 -----DCHADIIWQSLINKRYLQLYPHLLTBEDVD 301
D 270 -----DCHADIIWQSLINKRYLQLYPHLLTBEDVD 301
D 356 HYVGVLEHTKQYDIDMTQRAYVTKEDICHSSSCQTLDKNETDKNGFILTNDPLD 415
QY 302 ---YDNFGLSCGFDD---DDAHAHNG-----KFW-----IDLRKRYE 336
D 302 ---YDNFGLSCGFDD---DDAHAHNG-----KFW-----IDLRKRYE 336
D 416 GLVYDAGNQARGVYRGKGFNERHTKXRSIGFYRYENPKNSMPDLSLIDRQDLKLS 475
QY 337 LRAEWK---QPPPGFEALRVHLN-----RNDYHDEKAGD-AVENFNNQTNQARIE 384
D 337 LRAEWK---QPPPGFEALRVHLN-----RNDYHDEKAGD-AVENFNNQTNQARIE 384
D 476 SRIHWYCTDYPHVARCRASLDKPNVRYTEKNDYQERLNLQPNWEKTNLGFTHKVN 535
QY 385 LRHQPGRGLKSGWGVYLGKSSAL---SATSEAVKQPMLLDNKVKQHYSPFGVEQANWON 441
D 385 LRHQPGRGLKSGWGVYLGKSSAL---SATSEAVKQPMLLDNKVKQHYSPFGVEQANWON 441
D 536 I-----AAGFGTHRSTLQGDLYAEVTLPPYTEKV-----YGEDNKVKQN 577
QY 442 FTLEGGVVRVEKASIRYDIALIDRENVYKOP-LPDL-----GAHRQ-----TA 484
D 442 FTLEGGVVRVEKASIRYDIALIDRENVYKOP-LPDL-----GAHRQ-----TA 484
D 578 PTAE-----EKEK---LOYGNGSYDKPRVYRKNTPELKTVCNMGNETAGNDCSPVITG 630
QY 485 RSPALS-----GNWYF-----TFQHK 500
D 485 RSPALS-----GNWYF-----TFQHK 500
D 631 RQVYLALRNHIAFGWADLGLGVRYDNHTFRSNDPWTGKGNHNSWAGVSLKTRHFV 690
QY 501 LSLTASHOERLPSTOEY-----AHGK-HVATNTFEVGNKHLNKRNSNIELALGYEGDR 554
D 501 LSLTASHOERLPSTOEY-----AHGK-HVATNTFEVGNKHLNKRNSNIELALGYEGDR 554
D 691 VSYRVSSGFRVPFAYELLYGVGTGASGKNDPLTQKEFLSRPLKSEKAFNQIEGLAVQGD 750
QY 555 WQYNLALYRNFRNGYIYAQTLLNDGPKSIEDDSEMLKVRNOSGADFYGAE----- 606
D 555 WQYNLALYRNFRNGYIYAQTLLNDGPKSIEDDSEMLKVRNOSGADFYGAE----- 606
D 751 GVIETSPFQNNYKLL-----ARADKYVGLGVV-----TFYNTQDVKLGNGIN 794
QY 607 --GRIYFK-----PTRY-----RIGVSGDYVRGLKNL--PSL-----PGREDAYGNR 646

```

```

RESULT 15
US-10-238-075-1077
; Sequence 1077, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1077
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-10-238-075-1077

```

```

Query Match      3.9%; Score 155.5; DB 12; Length 682;
Best Local Similarity 18.9%; Pred. No. 1.5e-05;
Matches 143; Conservative 103; Mismatches 222; Indels 289; Gaps 40;

QY 51 LHTSTASDKIISGDTLROKAVNLGDALDGVPGIHASQYGGGASAP---VIRGQTRRIKVL 108
D 51 LHTSTASDKIISGDTLROKAVNLGDALDGVPGIHASQYGGGASAP---VIRGQTRRIKVL 108
D 44 LNAFASVSVITSEQLQKKPV---DLVADKDVESIGITGNEKDPDISIRGLSGDYTLIL 101
QY 109 -----NHGETGMADFPD-HAIVMDTALSOQVEILRGPVTLIYSSGNVAGLV 156
D 109 -----NHGETGMADFPD-HAIVMDTALSOQVEILRGPVTLIYSSGNVAGLV 156
D 102 VDRRQSGRESRPNSSGGPFAEIPPEAI-----ERIEVIRGPMSSLYGSDAIGVY 154
QY 157 DVADGKIPEK-----MPENG-----VSGELGRLSSG-NLEKL 188
D 157 DVADGKIPEK-----MPENG-----VSGELGRLSSG-NLEKL 188
D 155 NIITKPVNNQWDGVGLGGIIQEHGKFGNSTNDFYLSGGLIKDKLGLQLYGMNRYKE 214
QY 189 TSGGINIG-----LGNFVLHT-----EGL---YRKSQDYA 216
D 189 TSGGINIG-----LGNFVLHT-----EGL---YRKSQDYA 216
D 215 DS---ISQGPAPKKNKNTATLQFTPTESQKFVEYKGNQVHTLTPGESLDATWTRG--- 269
QY 217 VPRYRNKLPSPRPFANGQHRVGLW-----RKFRYRTSRRDQYCL 262
D 217 VPRYRNKLPSPRPFANGQHRVGLW-----RKFRYRTSRRDQYCL 262
D 270 -----NLKQ-PNSKRETHNSRSHWAAWNAQAQGEILHPEIAVYQEVKSVKSKDKYN- 322
QY 263 PAHSHYDDCHADIIWQ-----KSLINKRYLQLYPHLLTBEDVDYDNPGISCGFHDD 315
D 263 PAHSHYDDCHADIIWQ-----KSLINKRYLQLYPHLLTBEDVDYDNPGISCGFHDD 315
D 323 -----HWDLVESRKPEITWTIIDAQVTAFLP-----ENV-----LTIG----- 356
QY 316 DAHAHAHNGKPIDLRNKRYELRABWKOPFPQGFALRVHLNNDYHDEKAGDAVENFVN 375
D 316 DAHAHAHNGKPIDLRNKRYELRABWKOPFPQGFALRVHLNNDYHDEKAGDAVENFVN 375
D 357 -----GQFHAELRDSATGKKT 374
QY 376 NOTONARIELRHQPIGRGLKSGWGVYLGKSSALSATSEAVKQPMLLDNKVKQHYSPFGVE 435
D 376 NOTONARIELRHQPIGRGLKSGWGVYLGKSSALSATSEAVKQPMLLDNKVKQHYSPFGVE 435
D 375 TETQSVSKQKAVFIEN-----EYATDLSALATG-----GLRLN-----HEIYG--- 414

```

QY 436 QANK-----DNFTLEGGVVEKOKASIRYDKALIDRENYKQPLDGAHQDTA 484
 Db 415 -SWNPLVAVYMLTDLNLKGGIAKAFRAPS-----EVSFGTITQGG 460
 QY 485 RSPALSGNWYFTPO---HKLSTASHQERLPSTQBLAHGKHVANTFEVGNK-----534
 Db 461 ASI-MYGNRDLKPETSVEBIEGIYNSDGSASATLFTNDFKNLTSYDICTKDPVTGL 519
 QY 535 -----HLNERSNNTIELAGYE-GDRQYNLALYRNFGNYIAYOTLANDGRGPKSIEDD 587
 Db 520 NTFYDNGANIRGVELAQIPVYDKHVS-----ANY-----TFDSR--RKSDDE 565
 QY 588 SEMKLVRYNOSGADFYGAEGEIVFKPTPRYRIG--VSGDYVRGRKLKPLSPGREDAYGN 645
 Db 566 S-----LNGKSLKGE-----LERTPRHAANKLEWDYTD-ITFYSSL-----NYTCK 608
 QY 646 RPFTAQDDQAPRPAARLGFHLKASLTDRIDANLDY 682
 Db 609 QIWAQA--RNGAKVPRVRNGF-----TSMIDIGLNY 636
 RESULT 16
 US-10-043-344-94
 ; Sequence 94, Application US/10043344
 ; Publication No. US2003008086A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Harkness, Robin E.
 ; APPLICANT: Schryvers, Anthony B.
 ; APPLICANT: Chong, Pelle
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Murdin, Andrew D.
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
 ; FILE REFERENCE: 1038-1221 MIS
 ; CURRENT APPLICATION NUMBER: US/10/043,344
 ; CURRENT FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: 08/649,518
 ; PRIOR FILING DATE: 1996-05-17
 ; NUMBER OF SEQ ID NOS: 160
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 94
 ; LENGTH: 908
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-043-344-94

Query Match 3.8%; Score 155; DB 15; Length 908;
 Best Local Similarity 19.3%; Pred. No. 2.7e-05;
 Matches 187; Conservative 122; Mismatches 360; Indels 300; Gaps 47;
 QY 14 LLINTPLAQAHETEQSVGLTETVWVGKSRPRATSGLLHTSTASDKIISGDTL-ROKAVN 72
 Db 16 LMTALPYAENVQAEQAEKQOLDTIOYKAKQKTRDNEVTGLKLVKSSDTLSKEQVLN 75
 QY 73 LGDALDGVPGIHASQYGGASAPV-IRGQTRRIKV-----LNHHG 112
 Db 76 IRDLTRVDPGIADVQEGGASGYSIRGMDKNRVSALTVDGVSQISQSYTAQAALGGTTRTAG 135
 QY 113 ETGMDAFSPHAIMVDTALTSQQEILRGPTLYSSGNVAGLV-----VADGK- 162
 Db 136 SSGALNEIYENV-----KAVEISKGSNSBYGNGALAGSVAFTKTAADIIIGGKQ 187
 QY 163 --IPKMPENG-----VSGELGLRLSSGNLEKL-----TSGGINIGLGRNFV 202
 Db 188 WGIQSKTAYSGKDHALTQSLALAGSGGAELLYTKRGREIHAHKDAGKVGQSFNRLV 247
 QY 203 LHTGLYKSG-----DYAVPR-----YENLKLRLP 227
 Db 248 LDED---KKEGGSQRYFIVEECHNGYACKNKLKEDASVYDKRTVSTQDYTGSRNLL 304
 QY 228 DSPRRFANGQHRAVLGWR--KRFFY-----RRTYS--DRRQYGLPAH--SHEY-----269

Db 305 ANPLEYSSQWMLFRPGWHLDNRHVGVAVLERTQOTFTDRM-TVPAYFTSEDDYVPGSLKG 363
 QY 270 -----DOCHAD-----IIMOKSLINKYLOLYPHLLTEED-----299
 Db 364 LKYSGDKNKABRLFVQGGSTLQIGYGTGVFYDERHTKRYGYEYVYHNADKOTWADYA 423
 QY 300 -VDYDNPGLS-----CGFHDDDDAHAAHANGKPMIDLNRK--YE-----LRAEWK 342
 Db 424 RLSYDRQIGIDLNRLLQOQTHCS-HDGSCKNCRPDGKPKVSYFKSDRMITYEESRNLFOAVFK 482
 QY 343 QPPP-----GFEALRVHLNRNDYH-HDEKAGDAV-----ENFENNQTON-----380
 Db 483 KAFDTAKIRHNSLNLGYDRFKSQLSHSDYTLQNAQVAYDLITPKKPPFPNGSKNDPVKV 542
 QY 381 -ARIELHQPTIGRLKSGWQVYLQKSSALSATSEAVKQPMLLDNKVQHYVFFGVBEQAN 438
 Db 543 SIGKTTVNTSPICRGN-----TYTDCPTRNIGGNGYAAV 579
 QY 439 WDNFTLEGGVREKOKASIRYDKALIDRENYKQPLDGAHQROTARSFALSGM-----492
 Db 580 QDNVELG---RWADVAGIRYDY---RSTHSEDKSVSTGTHRN-----LSWAGVVLK 626
 QY 493 ---WYFTPQHKLSLTASHOERLPSTQELYAHGKHVATNTFEVGNKHLNKRSSNNIELALG 549
 Db 627 PFTM-----DLTPASTGFRLPFAEYNGWRAGESLTLB-----LXPEKSFREAGIV 676
 QY 550 YEGDRWQYNLALYRNFGNYI---YAQTLNDR-----GPKSIEDSEMKLVRYNQS-AD 601
 Db 677 PKGDFGNLEASVFNAYRDLIAFGYETRTQNGQTSASGDPGYRNAQARIAGINILGKID 736
 QY 602 FYGAEGEIVFKPTPRYRIGVSGDYVRGRKLKPLSPGREDAYGNRPFTAQ--DQONAPRV 659
 Db 737 WHGVWGL---PDGLYS---TLAYNRIRVKD-----ADTRADRTFVTSVLFDAVQPSR 783
 QY 660 PAARLGF-----HLKASLTDR-----IDANLDYRVFAONKLARYETRT 698
 Db 784 YVLGLGYDHPDGIWGINTMTFTYSKASVDELLGSQALLNGNANAKAASRRTPWYVTDV 843
 QY 699 PG-----HHMLNLCANYRNRTRYGEW-NWYVKADNLLNQ--SVYAHSSFLSDTPMGERS 749
 Db 844 SGYNYIKKHLTLRAGVYNLLNYRVYVWENVRQTAGGAYNQHKNVGVNRYAAP-----GRN 899
 QY 750 FTGQVNVKF 758
 Db 900 YTFLEMKF 908
 RESULT 17
 US-09-815-242-5085
 ; Sequence 5085, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haseibeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Cart, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5085
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5085

Query Match          3.8%; Score 154; DB 9; Length 813;
Best Local Similarity 20.0%; Pred. No. 2.8e-05;
Matches 162; Conservative 101; Mismatches 265; Indels 284; Gaps 43;

QY 47 TSGLLHTSTASDKIISG---DTRQKAVNLGDAL-DGVPGIHAS-----QYGGGASAPVI 97
Db 131 TSSDLVSATRQSTVIEHAQLBELRQGSDSLATVLAKAVPGMSDSSRTITEY-----181
QY 98 RGQT--GRRIKVLNHGSETGMADFPDHALMVDLTALSOQVEILRGPTVLLYSXGNVAGL 155
Db 182 -GOTLRGKSMVLWVDGVPNTNRSSRNLA-NIDPALIERIEVIRGS-SAIYSGATGGI 238
QY 156 VDV----ADGKIPKMPENGVSGBGL-LRSS-----GNLEKLTSGGINIGLGRNF---VL 203
Db 239 ISITRPAAG--ENRAETRLSATSPLTRGSDGLGGFOQYFAGSLG-ALDYSFDGTR 294
QY 204 HTEGLYRKSGDYAVPRYNLKRLPDSPRRFANGQHR--AVLGWRKRYRRTYSDRRDQYG 261
Db 295 HVGASYDAHGRIAPE-----ESQGLDFSNYINIGGLGLR-----331
QY 262 LPAHSHEVDCHADIIWOKSLINKRYLQLYPHLLTEEDVDYDN-----PGLSCGFHD 313
Db 332 -----IDE-----NQRVOLALSHYDARQOTDYATDPRVARLPFG-----365
QY 314 DDDAHAAHNGKPIDLANKYEILRAEWKQFPFGFEALRVHLNRNDVHHD-----363
Db 366 --SVPAKAIKLE-IDEQNRINTIANLE--YENLDILGSRLSAQLYRDFYTRFTPFDA 420
QY 364 ---EKAGDAVENFNNTQNTARTELHROPTGLRKGSS--WGVOLGOKS-----S 407
Db 421 RAVSTRGQNVQIMQNSVFQSRILTLR-TPLGESGNTLWGGDYNGQERSDMLDVEDPA 479
QY 408 ALSATSEAV-----KQPMLLDNKVQHYSFFGVEQANWD-NFTLEGGVVRKQKA-----455
Db 480 AYDASGGLVFDKIGKLTYPPLRTSRAGAFQAQLQHRFDEHWSIDGGLRYEYSTAEFDFFI 539
QY 456 -----SIRYDKALID-----RENY--YKQ--PLPDLGAHROTAR 485
Db 540 PLSKSAASPVTVKGGDLDYDAVLSNLGIVYSPVAGQEIYASFQGGFQDPDVGQLRNAR 599
QY 486 SPALSGNMYFTPOHKLSTASHQERLPSTQBELYAHGKHVATNTFEVGNKHLNKNERSNIE 545
Db 600 -----RGFDIGSSNLEPVKTNYYE 618
QY 546 LA-LGYEGDRWOYNLALY--BNRFGNYIYAOTLNDGRGPKSIEDDSEMKLVRYNQSGADF 602
Db 619 LGWRGAIQNTLGLSLALFTYTSKLGD---VQSFNNG-----LILTRTKER---I 661
QY 603 YGAGEEI-YFKPTPRYRIGVSDYVVRGLKNLPSLPGREDAYGNRPFIADQDQNAVRPA 661
Db 662 YGVEASADWLSDDEVWAGGAGSATWNRGREK-----PDGKD-----WQDMTGYRVP 707
QY 662 ARLGFH-----IKASLTDRIDANLDYRVPFQAKNLARYET-----RTGHH 702
Db 708 LKLTAYLQYKPDADWNRLQNTFFDFSKDYRLDGVESFGRRQVSTYTTVDLVQSYRITPDD 767
QY 703 MLNLGANYRNRTRYGEWNWYKADNLLNQSVY 734
Db 768 QLSLG-----IQNLFNRDIY 782
```

Db 535 GYEKQBYLXSKPVDVFGQDHCYKSSNSYSDCKVRILKGNVYFAARNMNLGKYIDL 594
Qy 478 G-----RSPALSGNMYFTPOHKLST--ASHQERLPSTQELIYA 519
Db 595 GLGIRDVSRKTANESTISVGKFKNSWNTGIVIKPTWELDLXYLSTGFRNPSFAEMY- 653
Qy 520 HGKRVATNTFVGNKHLNKRNNIELALGYEGDRQWYNALYRNFNGYI-YAQTLLN-- 576
Db 654 -GWRYGNNSDVYVGFKPKETSRNOBFGLGKDFGNIHFSNAYRNLAFAPELSKN 712
Qy 577 -DGRPKSIEDSEMKLVRYN--OSGADFYGAEGEYF----- 611
Db 713 GTTGKNGYHNAQNAKLGVNITAGLDENGLMKLIPYCWYATFAYNRVKVKKOQKINAGL 772
Qy 612 -----KTPRYRIGVSGD-----YVRGLKMLPSLPQREDAYGNRPFI 649
Db 773 ASVSSYLDAIQPS-RYIILGLGYDHPNSNTWGTQSKAKSQNELIGOR-ALGN----- 826
Qy 650 AODQONAPRVPARLGFHLKASLTDRIDANLDYRVAQNKLYRTPGHHMLNLGAN 709
Db 827 --NSRVKSTRKLTAWHI-----LDVSGYYMANK-----NIMRLGIY 863
Qy 710 YRNTRYGENWVVK-ADNLLN--QSVYAHSSFLSDTPQMGSRFTGCVNVKP 758
Db 864 NLFNRYVTWEAVRQTAQGVNQHONVGSYTRYAAS-----GRNYTLTLEMKF 911

RESULT 19
US-10-240-218-6
; Sequence 6, Application US/10240218
; Publication No. US20030186848A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: Goringe, Andrew Richard
; APPLICANT: Hudson, Michael John
; APPLICANT: Matheson, Mary Anne
; APPLICANT: Robinson, Andrew
; APPLICANT: West, David McKay
; TITLE OF INVENTION: Recombinant iron uptake proteins
; FILE REFERENCE: GMS/22057
; CURRENT FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: US/10/240,218
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-240-218-6

Query Match 3.8%, Score 151.5; DB 12; Length 915;
Best Local Similarity 19.0%; Pred. No. 6e-05;
Matches 191; Conservative 131; Mismatches 345; Indels 339; Gaps 50;
Qy 1 MAQTLKPIVLISLIINTPLLAQAHETE---QSVGLETVTVVVGKS---RPRATSGLLH 52
Db 1 MQQHLFRFNILCSLMTALPAYAENVQAQAEKQDLTIQVAKKQKTRDNEVTGL-- 58
Qy 53 TSTASDKIL-SGDTL-RQKAVNLGDALDVPFGIHASQYGGGAPV-IRGQTGRRIKV-- 107
Db 59 -----GKLVKSSDLSKEQVNLNRLDLYRDPGIAVQVQGRGASGSYIRGMKNRVSALT 113
Qy 108 -----LNHHGETGDMADFPDHAIMVDTALSQVEILLRGPVTLTYSSG 150
Db 114 DVSVQIQSYTAQAALGGTRTAGSGGAINETEVENV-----XAVEISKGSNVEQSG 165
Qy 151 NVAGLV----DVADGKIPEK-----MPENGVSGLRLSSGNLEKLTSGGINIGLKNF 201
Db 166 ALAGSVAFQTKTADDDVIGEGRQWGIQSKTAYSGK-----NRLTQSIALAG-RIGGAEL 219

Qy 202 VLHTEGLYKSGDY-----AVPRYRMLKELPDS----- 229
Db 220 LIHTG--RRAGEIRAHEDAGRGVGSFNLVPEVDESSNAYFIVKEECXGNGSYETCKANP 276
Qy 230 -----PRRF-----ANGQRAVL-----GWR-----KRFY-----RR 251
Db 277 KXDVGVGDERQTVSTRDTGPNRFLADPLSYESRSWLFPRPGFRFENKRHYIGGILEHTQQ 336
Qy 252 TYSDR-----RDQYQ-LP-----AHSHEYDDCHAD----- 275
Db 337 TFDTRDMTVPAFUTKAVFDANKQAGSLPQNGKYAGNHKHYGGLFTNGENGALVGAEGYTG 396
Qy 276 IIWOKSLINKRYLQLYPHLLTBE-----VDYDNPGLSCGFH-----DDDDAHAAH 320
Db 397 VFDETHTRKSYGLEVYVYTNADKDTWADYARLSYDQGVGLDNHFQOHCSDGSKYCR 456
Qy 321 AHNGKPIDLRVRY-----ELRAEWKQPP-----GFEALRVHLNENDVH 361
Db 457 PSADKPFYKSRVYIGSHRLLQAFKFSFDTAKIRHNLNVNGLDFRFGNLRHQDY 516
Qy 362 HDEKAGDAVENFFNNTQNAIRLHQ-----PICRLKSGWGVQYLGQ-----KSSAL 409
Db 517 Y-QHANFAYSSNTFPQNGKKISPNGSETSPYVWTIGR-----GNVVTGOICRLGNNTYT 570
Qy 410 SATSEAVKQMLLDNKVQHYSPFGVQANWDNFTLEGGVVEKQKASIRYDKALIDRENY 469
Db 571 DCTPRSI-----NGKSYIA-----AVRDNVRLG-----RWADVAGAGLYD----- 605
Qy 470 YKQPLDPLGA-HRQTARSPALSGNMYFTPOHKLST--ASHQERLPSTOELVYAHGKHVAT 526
Db 606 YRSTHSDGVSSTGTHRTLSWAGIVLKPDTMDLDLYRTSTGFLPSPFAEMYGWRAGVQS 665
Qy 527 NTFEVGNKHLNKRNNIELALGYEGDRQWYNALYRNFNGYI---YAQTLDG---R 579
Db 666 KAVKI-----DPEKSFNKEAGIVFKGDPGNLEASWENFNNAYRDLIVRGYZAQIKDGKEAK 720
Qy 580 GPKSIEDDSEMKLVRYNOSG-----ADFYGAEG 607
Db 721 GDPALVNAQARITGINILKIDWNGVDKLPBGWYSTFAYNRVVRDIKKRADRTDIQS 780
Qy 608 EIFYKPTP-RYRIGVSGDYVRG-----RLKNLPSLPQREDAYGNRPFIADQDN 655
Db 781 HLFDAIQPSRYVWGLGYDQPEGKMGVNGMLTYSKAKEITEL-----LGSRALINGNSRN 834
Qy 656 APRVPAARLGFHLKASLTDRIDANLDYRVAQNKLYRTPGHHMLNLGANYRNT 715
Db 835 TKATARTRPWYI-----VDVS-GYITV-----KKHFTLRAGVYNNLLNYR 873
Qy 716 YGEW-NWYVKADNLLNQ--SVYAHSSFLSDTPQMGSRFTGCVNVKF 758
Db 874 YVTWENVRQTAGGAVNQHKNVGVYNYAAP-----GRNYTFSLEMKF 915

RESULT 20
US-09-332-226-2
; Sequence 2, Application US/09332226
; Patent No. US2002025318A1
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/332,226
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/363,124
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/973,336
 FILING DATE: 05-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/572,187
 FILING DATE: 23-AUG-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Gallagher, Thomas C.
 REGISTRATION NUMBER: 37,066
 REFERENCE/DOCKET NUMBER: SPA-1-PDC
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 645-1405
 TELEFAX: (212) 645-2054
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 915 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-332-226-2

Query Match 3.7%; Score 150.5; DB 9; Length 915;
 Best Local Similarity 18.4%; Pred. No. 7.5e-05;
 Matches 186; Conservative 132; Mismatches 344; Indels 349; Gaps 48;

QY 1 MAQTLKPIVLSILLINTEPLLAQAHETE---QSVGLETVTVVGS---RPRATSGLLH 52
 DB 1 MQQHLFRNLICLSLMTALPAYENVQAGQAEKQDITQVAKKQKTRDNEVTGL-- 58
 QY 53 TSTASDKII--SGDTL-RQKAVNLGDALDGVPGIHASQYGGGASAPV--IRGQTGRIRKV-- 107
 DB 59 -----GKLVKTADTLSKEQVLDIRLTRYDPGIAVVEQGRGASSGYSGIRGMDKNRVS LTV 113
 QY 108 -----LNHHGETGDMADFPDHAIMVDLTALSOQVEILRGPVTLIYSSG 150
 DB 114 DGLAQIOSYTAQAALGGTTRTAGSGAINEIEYENV-----KAVEISKNSNVEQSG 165
 QY 151 NVAGLV-----DVADGKPIKMPENGVSGLRLSSGNLEKLTSGGINIGLKNFV L-- 203
 DB 166 ALAGSVAFQTKTADDOVI-----GEGRWGIGQ-----SKTAYSGKNRGLTQSIALAGR 212
 QY 204 --HTEGLYKSGDYA-----VPRYENLKRLPDSEPR-----RFANGQH--- 238
 DB 213 IGGAEALLIRTGRIHAGEITRAHEAAGRGVQSPNRLAPVDGSGKYAFIVEECKNGGHEK 272
 QY 239 -----RAVLG-----W-----PKRFY-----R 250
 DB 273 KANPKDVVGEDKQKTVSTRTDTGPNRFLADPLSVESRMLFRPGFRENKRYIGGILS 332
 QY 251 RTYS--DRDQ-----YGLPAHSHEYD-----D 271
 DB 333 RTQQTFFTRDMTPAFUTKAVFDANQKAGSLRGNGKYAGNHKYGGLFTSGENNAAPVGA 392
 QY 272 CHADIIWOKSLINKRYLQLYPHLLITEED-----VDYDNPCLSCGFH-----DDDD 316
 DB 393 YTGCVFYDETHKSYGLEVYVTNADKTWADYARLSYDRQIGLDNFHFOOTHCSADSD 452
 QY 317 AHAAHNGKPMIDLNKRY-----ELRAEWKQFPF-----GFEALRVHLNR 357
 DB 453 KYCRPSADKPSYKSDRVIYGESHKLLQAFAKKSFTAKIRHNLVSNLGYDRFGSLNRH 512
 QY 358 NDYHDEKAGDAVENFNQNTQNAIRIELRHOPIGRLKSGWGVYLGQKSSALSATSVA 417
 DB 513 QDIYY-----QSANRAISLTKPPQNNGKTSFNGREKPYVW-----SIGRNVVTR 559

QY 418 QPMLLDN-----KVQHYSPFGVQQAQWNDNFTLEGGVVRVEKQKASIRYDKALIDREN 468
 DB 560 QICLFGNNTYDCTPRSGNGSKSYAAVR---DNVRLG---RWADVAGAGLYD----- 605
 QY 469 YKQPLPDIGA-HRQTARSPALSGNWTFTPOHKLST--ASHQERLPSTOELVAHGKHA 525
 DB 606 -YRSTHSDGVSCTGTHRTLSWAGIVLKPADWLDLTYRTSTGFRFLSPFAEMYG----- 658
 QY 526 TMTFEVGNK-----HLNKERSNNIELALGYEGDRNQYNLALYRNRFPGNYI---YAQTLDNG 578
 DB 659 ---WRSQDKIKAVKIDPEKSPKEAGIVFKGDFGNLEASWFNNAAYRDLIVRGVEAQIKDG 715
 QY 579 R-----GPKSIEDDSEM-----KL-----VRYNQSADP 602
 DB 716 KEQVKNPAYLNAQASRITGINILKIDWNGVWDKPEGWYSTPAYNRVRDRIKKAADR 775
 QY 603 YGAEGEY--FKPT-----PRYRIGVSGDYVRGRLNKLPFLPREDAYQNRPFIA 650
 DB 776 TDIQSHLFDATQPSRYVVGSGYDQPEGKMGVNGMLTVSKAKEITEL-----LGSRALLN 829
 QY 651 QDDQNAQPRVPAARLGFHLKASLTDRIDANLDYVYFAQNKARLYETRTPGHMLNLGANY 710
 DB 830 GNSRNTKATARTRPWYI-----VDVS-GYTYV-----KHFTLRAGVYN 868
 QY 711 RNRTRYGEW-NWYVKADNLLNQ--SVYAHSSFLSDTPQMGSRFTGCVNVKPF 758
 DB 869 LLNHHYVTVWENVRQTAAGAVNQHKVNGVYNYAAP---GRNYTFSLEMKP 915

RESULT 21

US-10-043-344-96
 ; Sequence 96, Application US/10043344
 ; Publication No. US20030080806A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loomore, Sheena M.
 ; APPLICANT: Harkness, Robin E.
 ; APPLICANT: Schryvers, Anthony B.
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Murdin, Andrew D.
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
 ; FILE REFERENCE: 1038-1221 MIS
 ; CURRENT APPLICATION NUMBER: US/10/043,344
 ; CURRENT FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: 08/649,518
 ; PRIOR FILING DATE: 1996-05-17
 ; NUMBER OF SEQ ID NOS: 160
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 96
 ; LENGTH: 915
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-043-344-96

Query Match 3.7%; Score 150.5; DB 15; Length 915;
 Best Local Similarity 18.4%; Pred. No. 7.5e-05;
 Matches 186; Conservative 132; Mismatches 344; Indels 349; Gaps 48;

QY 1 MAQTLKPIVLSILLINTEPLLAQAHETE---QSVGLETVTVVGS---RPRATSGLLH 52
 DB 1 MQQHLFRNLICLSLMTALPAYENVQAGQAEKQDITQVAKKQKTRDNEVTGL-- 58
 QY 53 TSTASDKII--SGDTL-RQKAVNLGDALDGVPGIHASQYGGGASAPV--IRGQTGRIRKV-- 107
 DB 59 -----GKLVKTADTLSKEQVLDIRLTRYDPGIAVVEQGRGASSGYSGIRGMDKNRVS LTV 113
 QY 108 -----LNHHGETGDMADFPDHAIMVDLTALSOQVEILRGPVTLIYSSG 150
 DB 114 DGLAQIOSYTAQAALGGTTRTAGSGAINEIEYENV-----KAVEISKNSNVEQSG 165
 QY 151 NVAGLV-----DVADGKPIKMPENGVSGLRLSSGNLEKLTSGGINIGLKNFV L-- 203

384 ELRHQPIGRKSGWVQYLGQKSSALSATSEAVKQPMLLDNKVQHVSPFGVQANWDF 443
 543 SI-----GKTTVNTSPFGGNNYTDCTPRNIGGNGYAAVQDNVR 584
 444 LEGGVRVEKOKASIRYDKALIDRENYKQPLDLAGHROTARSFALSNG-----WY 494
 585 LG---RWADVAGIRYD---RSTHSEKSVSTGTHRN-----LSWAGVVLKPFTHM 631
 495 FTPQHKLSLASHQERLPSTQELAHGKHVATNTFFVGNKHLNKRSSNNIELALGYEGDR 554
 632 -----DLTYRASTGFRLLPSFAEMYGWRAGBSLKTLD-----LKPEKSFNREAGIVKGGDF 681
 555 WQYNALAYRNFNGYI---YAQTLNDRGPKSIEDDSMKLVRYNQSGADFYGAECEIYP 611
 682 GNEASFNNAYRDLIAFGYETRTQNGQ-----TSASGD----- 715
 612 KPTPRYRIG-----VSGDYVRGRILK-----NLPSLPGREDAYGNRPFFIAQDDQNA 658
 716 ---PGYRNGPKCTVAGINILGKIDHWGVWGGLPDGLYSTLAY-NRIKVKADADIRADRTF 771
 659 -----VPAAR-----LGF-----HLKASLTDR-----IDANLDYVRV 685
 772 VTSYLFDAVOPSRVYLGVDHPDGIGWINTMTFTYSKAKSVDELGSQALLNGNANAKKA 831
 686 FAQNKLARVETRTPG-----HHMLNLGANYRNRTRYGEW-NWYVKADNLNQ---SVYAH 736
 832 ASRETRPWYTDVSGYNYIKKHLTLRAGVYNLLNRYVVTWENVROTAGGAVNQHKNGVY 891
 737 SSFLSDTPQMRSGFTGGVUNKF 758
 892 NRYAAP-----GRNYTFSLEMKF 909

RESULT 24
 US-10-043-344-95
 ; Sequence 95, Application US/10043344
 ; Publication No. US2003008086A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Harkness, Robin E.
 ; APPLICANT: Schryvers, Anthony B.
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Murdin, Andrew D.
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
 ; FILE REFERENCE: 1038-1221 MIS
 ; CURRENT APPLICATION NUMBER: US/10/043,344
 ; PRIORITY FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: 08/649,518
 ; PRIOR FILING DATE: 1996-05-17
 ; NUMBER OF SEQ ID NOS: 160
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 95
 ; LENGTH: 911
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 US-10-043-344-95

Query Match 3.6%; Score 144.5; DB 15; Length 911;
 Best Local Similarity 19.0%; Pred. No. 0.00028;
 Matches 191; Conservative 135; Mismatches 337; Indels 343; Gaps 52;

QY 1 MAQTLKPIVLSILINPLLAQAHEE---QSVGLETVTVVGKS---RPRATSGLLH 52
 DB 1 MQQHLFRINILCLSMALPAYENVQAGQAEKQDITQVAKKQKTRDNEVTGL-- 58
 QY 53 TSTASDKII-SGDTL-RQKAVNLGALDGVPGIHAQYGGGASAPV-IRGOTGRRIKV-- 107
 DB 59 -----GKLVKTADTLSEQVLDIRLTRYDPGIAVVEQGGAGSGYSIRGMDKNVSLTV 113
 QY 108 -----LNHHGETGDMADFDPDHAINVDITALSQQVEILRGFTVLLYSSG 150

Db 114 DGLAQISYTAQAALGTRTAGSSGAINEIYENV-----KAVEISKSGNSVEQSG 165
 QY 151 NVAGLV-----DVAGCKIPEK-----MPENGVSGBGLRLSSGNLEKLTSGGINIGLKNF 201
 Db 166 ALAGSVAFQTKTADDDVIGEGRQMGIOKSTAYSGK-----NRGLTQSIALAG-RIGGAEL 219
 QY 202 VLHTEGLYKSGDY-----AVPRYNLKRLLPDS----- 229
 Db 220 LIHTG-----RRAGEI RAHEDAGRGVQSFNRLVPVEDSSEYAYFIVEDCEGKNYETCKSKP 276
 QY 230 -----PRRF-----ANGQHRVAVL-----GWR-----KRFY-----RR 251
 Db 277 KQDVVGKDERQTVSTRODTGPNRFLADPLSYESSRWLFRPGFRPENKRHYTIGGILEHTQQ 336
 QY 252 TYSDBRD-----QYG-LP-----AHSHEYDDCHAD----- 275
 Db 337 TF-DTRDMTVPAFLTKAVFDANSQOAGSLPONGKYAGNHKYGGLFTNGENGALVGAEGYT 395
 QY 276 -IIMOKSLINKRYLQVPHLLTEED-----VDYDNPLGSCGFH-----DDDDAHA 319
 Db 396 GVFYDETHTKSRYGLEYVVTNADKDTWADYARLSYDRQIGLDNHFQOQTHCSADGSKYC 455
 QY 320 HAHNGKPMIDLNKRY-----ELRAEWKQFPF-----GFEALRVHLNRNDY 360
 Db 456 RPSADKPFYSYKSDRVIYGESHRLLQAAFKSFDTKIRHNLVNLGDFRDSNLRHODY 515
 QY 361 HHDEKAGDAVENFNQOTONARIELRHOPIG-RLKGSW-----GVQYLQ-----KSSAL 409
 Db 516 YY-QHANRAYSKTPPKTAN-----PNGDKSPYVWSIGGNNVVTGQICLFGNNYIT 566
 QY 410 SATSEAVKQPMLLDNKVQHVSPFGVQANWDFTEGGVVRVEKQKASTRYDKALIDRENY 469
 Db 567 DCTPRSI-----NGKSYA-----AVRDNVRLG-----RWADVAGALRYD----- 601
 QY 470 YKQPLDLGA-HRQTARSGFALSNGWYFTPOHKLSTL--ASHQERLPSTQELYAHGKHVAT 526
 Db 602 YRSTHSDGSGVSTGTHRTLSWAGIVLKPADWLDITYTSTGFRLPSPAEMYGWRSGVQS 661
 QY 527 NTFEYGNKHLNKRSSNNIELALGYEGDRQWYNLALYRNFNGYIY-----AQTNL---DGR 579
 Db 662 KAVKI-----DPEKSFKEAGIVPKGDFGNLEASFNNAYRDLIVRGYEAQIKNGKEBAK 716
 QY 580 GPKSIEDDSSEMKLVRYNQSG-----ADFYGAG 607
 Db 717 GDPAYLNAQSARITGINILGIDMNGVMDKLPPEGWYSTFAYNRVHVRDIKKRADTDIQS 776
 QY 608 EYFVKPTP-RYRIGVSGDYVRG-----RLKNLPSLPGREDAYGNRPFFIAQDDQN 655
 Db 777 HLFDAIQPSRYVVGGLGYDQPEGKMGVNGMLTVSKAKEITEL-----LGSALLNGSRN 830
 QY 656 APRVPAARLGFHLKASLTDRIDANLDYRVPFAQNKLAARYETPTPGHMLNLGANYRNRTR 715
 Db 831 TKATARTPTWYI-----VDVS-GYVTI-----KKHFTLRAGVYNLLNRY 869
 QY 716 YGEW-NWYVKADNLNQ--SVYAHSGFSLSDTPQMRSGFTGGVUNKF 758
 Db 870 YVTWENVROTAGGAVNQHKNGVYNRYAAP-----GRNYTFSLEMKF 911

RESULT 24
 US-09-815-242-10132
 ; Sequence 10132, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10132
; LENGTH: 729
; TYPE: PRF
; ORGANISM: Escherichia coli
; US-09-815-242-10132

Query Match      3.5%; Score 141.5; DB 9; Length 729;
Best Local Similarity 20.0%; Pred. No. 0.00038;
Matches 151; Conservative 93; Mismatches 275; Indels 235; Gaps 35;

QY 3 QTTLKPIVLS--ILLINPPLA-QHETQSV---GLETVVVVGKSRPRATS-----GLLHT 53
Db 13 QAITKPSLLAGGIALALLPSAFAPAETEETVIVEGSATAPDDGENDSVTSTAGTKMQ 72
QY 54 STASD-----KTIISGDTLR-QXAVNLGDALDGPCHASQYGGASAPVIRGQTRRIKV 107
Db 73 MTQRIPOSVTIVSQRMEDQQLQTLGEVMENTLGISKSQADSORALYSRG-----PQI 127
QY 108 LNHGEGTGMADSPDH-----AIMVDTALSOQVEILRGPTVLLYSSGNVAGLVDVA 159
Db 128 DNY-----MVDGIPTYESRWNLGDALSMALFERVEVRGATGLMTGTGNSPAAINNV 181
QY 160 DGIKEKMPENGVSGLGLRLSSNLEK-----LTSGGINIGLKNFVHLHTEGLVRK 211
Db 182 RKHATREFKGVDSAEY-----SWNKRYVADLQSPLETDG-----KIRARIVGGY-Q 229
QY 212 SGDYAVPRYRNLRPLDPSRRFANGCHRAVLGWRKRF-----YRRTYSDRDQVGLP--- 263
Db 230 NNDMSLDRY-----NSEKTFPGIVDADLGLDITLTSAGVEYQRIIDVNSPTWGLPRWN 282
QY 264 --AHSHEYDDCHA--DIITWQSLINKRYLIQLYPHL-----LTREDVDYDNPGLSG 309
Db 283 TGGSNSYDRASTAPDWAYNDKEINKVFWTLKQOFADTWQATLNAHSEVEFDSKMYV 342
QY 310 GFHDDDDAHAAHNGKPTWIDLRNKRYELRAEWKQFPFGFEALRVHLNRNDYHHDEKADA 369
Db 343 -----DAYVNKADGN-----LVGPSYNGPGFQ-----YVGGTWNSSGKRKYDA 381
QY 370 VENPF-----NNQTONARIELRHQIGRL-----KSGMGV 399
Db 382 LDLEADGYSYELFGROHNLNMFSGYSKQNNRYFSSWANIPPDIGSFYFNFGNFPQTDWSP 441
QY 400 QYLGQ-----KXSALSATSEAVKQPMILDNKVQHSFFGVEQANNDNFTLEGVVRVEKOK 454
Db 442 QSLAQDTHMKSLYATATVTLADPLHL-----ILGARYTNW----- 478
QY 455 ASTRYDKALIDRENYKQPLDPLGAHQRTARFALSNGNYFPTQHKLSLTASHOERLPST 514
Db 479 --RVDTLAYSMEKNHTTEYAGL-----VFDINDNW-----STYASVTSIFQFPQ 519
QY 515 QELIYAHGKHVATNTEFVGNGHNLKERSNNIELAL--GYEGDRWQYNLALRYRNFNGYIYA 572
Db 520 NORDSSGKYLAIPIT-----GNNYELGLKSDMWNRLTTTLAIFR----- 558

```

```

QY 573 QTINDGRGPKSLEDSEMKLVRYNOSGADFYGAEGEIVYEKPTPRYRIGVSGDYVRGLKN 632
Db 559 -----TEQDNVAQ-----STGTPIPSNGETAYK-----AVDGTVSGVEFE 595
QY 633 LPS--LPGREDAYGNRPFTIAQDDQ-----NAPR 658
Db 596 LNCATIDNWLQTLTGATRYIAEDNEGNAVNPMLPR 629

RESULT 25
US-09-881-752A-334
; Sequence 334, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 791
; TYPE: PRF
; ORGANISM: Helicobacter pylori
; US-09-881-752A-334

```

```

Query Match      3.3%; Score 135; DB 10; Length 791;
Best Local Similarity 19.4%; Pred. No. 0.0018;
Matches 170; Conservative 112; Mismatches 352; Indels 242; Gaps 46;

QY 19 PLLAAQHETBQSVLETVTVVVGKSRPRATSGLLHTSTASDKIISGDTLRKAVNLGDALD 78
Db 16 PLLAETCYBERT-LNKVTTQAKR-----IFTYNEEFKVTISKELDQOSNEVDLFR 66
QY 79 GVPCHASQYGGGA---SAPVIRGQTRRIKV-----LNHGETGMADSPDH 124
Db 67 TNPDVNV---GGSGVMGQKIYVRGVEDRLRVTVGAAQNGNIYHHQGN----- 113
QY 125 AIMVDTALSOQVEILRGPTVLLYSSGNVAGLVDVADGKIPEKMPEN---GVSGELGLRLS 181
Db 114 --VIDPMLKSVEVTKGANASAGPGNAGIAGVIMETYGADFIPIRGKNAASGAVSFYTN 171
QY 182 SGNLEKLTSGGINIGLG-KNFVLTGELYRKSGDYAVRYRN---LKRLPDSRRFANGQ 237
Db 172 FGORETFRSAQNAHFIIAYTHQIFIVYRSGATAMKLNFNPTQADKEPGTP---SEQ 227
QY 238 HRAVLGWRKRYR-----TYSDRDQYGLPAHSHEYD---DCHADIIWQSLINKRYLQ 289
Db 228 NNALIRNNGYLSRDITLTFSWNMTNRNATRLPSRNAIGLAIYPCAPFSPDSSQGCNVLD 287
QY 290 LYPHLLTEEDVDYDNPGLSGCFHDDDAH-----AHAIN--GKFWID-----LRNKR 334
Db 288 SFTRYM-----YHSINSANNLSLQYKREAGNSFGDPRLDFTLTYSIRNAQ 332
QY 335 YE-----LRAEW-KQFPFGFEALRVHLNRNDYHHDEKADGAVENFNNQT 378
Db 333 FDLFDLPNGYVAKFPTSLASAMEKENYPCVEGAYCTPFSADV---DKPSSQPNRLFNNT 389
QY 379 QNARIELRHQPIGLKGS---WGVQY-----LGOKSSALS 410
Db 390 -GLNLKVAH-VIDEATDSLFEYFNQYONLSVFDPAIPKSELYRPNQYVTTDKQKQIACS 447
QY 411 ATSEAVKQPMILD-NKVQHSYFFGVEQANWDN---FTLEGGVVRVEKQKASIRYDKALIDR 466

```


Db 448 LVNNNDPFLCORGKANGNIYGYQANYSPHKIITFGAGVRWD---AYTLYDK---DW 501
 Qy 467 ENYKQPLDGLGAHROTARSEALSGNWFYTPQKLSLTASHOERLPSTQELYAHGKHVAT 526
 Db 502 NHRITO---GFSPSAALVLSPIEPLSLKITYSQ---VTRGVWPGDGVYMRQ 546
 Qy 527 NTEVEGNKHLNERSNNIELALGYEGDRWOYNLALYRNFRNGYI---YAQTLDNGRGPKSI 584
 Db 547 NDLRYA-KNIKPEVGSNAEFNIDYSSQYFSGRAAFAFYQALDNFISQYQNL----- 596
 Qy 585 EDSSEMKLVYNGSGADFFYCAE---GRIYFKPTPRYRIGVSGDY---VRGRLKNLPSLPGR 640
 Db 597 ---IVNLSQAIRIYGYEGGTFRYKGV-SLVNGVSRTPWTRGYL-----MA 640
 Qy 641 DAY-----GNRPPIAODDQNAVRPAPARLGHKLKASLTDRIDANLDY---YRVFAQNKIA 692
 Db 641 DSYELAASTGN-VFIILKDYTIPTK-----GINL---AWLSRFVTGLDYCGFDIYLFYGT 692
 Qy 693 RYETRTD-----GHEMLNIGANVRNTRYGEWNWYVK-----ADNLL 729
 Db 693 AEKPTITDLAKGS---QLGLVHMHPGYGVSNFVNNWSPKTSKRWKGLLSAVFNNVF 749
 Qy 730 NQ-----SVYHSSFLSDTPQGRSFT-CGVNVK 758
 Db 750 NKFYVDQTSFYVWSPMDPGTDAVKRAITAEFGFNARF 785

RESULT 26

US-09-815-242-5106
 ; Sequence 5106, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA 011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5106
 ; LENGTH: 815
 ; TYPE: PR1
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-815-242-5106

Query Match 3.1%; Score 127; DB 9; Length 815;
 Best Local Similarity 17.4%; Pred. No. 0.011;
 Matches 135; Conservative 95; Mismatches 242; Indels 302; Gaps 36;
 Qy 21 LAQAHETEQSGLETWT-----VVGKSRPRATSGLLHTST-----ASDKIISGDT 65

Db 117 VAEAAUSSVDLGATMTITSNQLGTTIEDSSYTPGTIATATRLVLTPTRETPQITVVTQON 176
 Qy 66 LROKAV-NILGDALDGVPGIHASQYGGASAPVIRGOTGRRRIKVLNHHG--ETGDMADPSP 122
 Db 177 MDDFGLNIDDDVHRHTPGITVSAYDTRNNYVARGFS---INNFOYDGIPTARNVGYSA 233
 Qy 123 DHAIMVDTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPEKMPENGVSSELGLRLSS 182
 Db 234 GNTLS-DMAIYDRVEVLKGAATGLTLAGSLGATINLI-----RKPTEFKGH--VELCA 285
 Qy 183 GNLEKLTSGGINICLGKNGFVLTGELYRKSGDYAVPRYRNKLKELPDSPRRFANGQRAYL 242
 Db 286 GSWDN-----YRSELDVSGP-----LTES-----GNVRGRAVA 313
 Qy 243 GWRKFRYRTYSRRDQ--YGLPAHSHVEDDCHADIIWQKSLINKLEYLOLYPHLLTEEDV 300
 Db 314 AYQDKHSGFMDHYERKTSVYVGI-----LEFD-----LNFDTMTVTGA 350
 Qy 301 DY-DNPGLSGCFHDDDDAHAAHNGKPWIDLRNRYELRAEWKQPPPGFEALRVHLNRND 359
 Db 351 DYQNDPKGSG-----MSGSPFLFDS---QGNRND 377
 Qy 360 YHDEKACADAVENFFNNQTONARIELRHQPIGRKSGWGVOLGOKSSALSATSSEAVKOP 419
 Db 378 -----VSRSFNN----- 384
 Qy 420 MLLDNKVQHYSGVEQANWDFTEGGVREVEKQASIRYDKALIDRE-NYKQPLDGLG 478
 Db 385 -----GAKWSSWEQYTRTVFANLEHNFANGVGVQLDHKNINGYHAPLG--- 428
 Qy 479 AHROTASFPALSGNNYTPQHKLSLRASHO--RELSTQELYAHGKHVATNTE-VGNKH 535
 Db 429 -----AIMGDWP--APDNSAKIVAOKYTGTGKNSLDIY-----LTGPFQFLGREH 472
 Qy 536 LNKERSNNIELALGYEGDRWQYNLALYRNFRNGYVIAQTLDNGRGPKSIEDDSEMKLVRY 595
 Db 473 -----ELVVGTSAS-----FSHW-----EGKSYWNLRY 496
 Qy 596 NQSGADFYGAGEIYFRTPRYRIGVSGDYVRGRLKNLPSLPGRDAYGNRPFIADQDN 655
 Db 497 DNTTDDFINWDGI--GKPD-----WGTPSQYIDDKTRQLGS-----YMTARENFVTD-- 542
 Qy 656 APRVPAARLGHKLKASLTDRIDANLDYRVFAQNKIARVETRTPGHMLNMGALYRNTR 715
 Db 543 -----LNLFLGGRVVD-----YRVTLNLP-----TIRSGRPIYVGVAYDLNDT 582
 Qy 716 YGEW-----NMVYKADNLLNQSVAHSSFLSDTPQMGSRFTGCVNVK 758
 Db 583 YSVVASYTDIETWPDQSWYRDSNNKLE-----PDEGQNYEIGIKGEY 624

RESULT 27

US-09-881-752A-362
 ; Sequence 362, Application US/09881752A
 ; Patent No. US20020115078A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kleanthous, Harold
 ; APPLICANT: Al-Garawi, Amal
 ; APPLICANT: Miller, Charles
 ; APPLICANT: Tomb, Jean-Francois
 ; APPLICANT: Oomen, Raymond P.
 ; TITLE OF INVENTION: Identification of Polynucleotides
 ; TITLE OF INVENTION: Encoding No. US20020115078A1
 ; FILE REFERENCE: 06132/041002
 ; CURRENT APPLICATION NUMBER: US/09/881,752A
 ; PRIOR FILING DATE: 2001-06-15
 ; PRIOR FILING DATE: 1997-04-01
 ; NUMBER OF SEQ ID NOS: 370
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 362

```

; LENGTH: 793
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-362

Query Match      3.1%; Score 125; DB 10; Length 793;
Best Local Similarity 19.8%; Pred. No. 0.017;
Matches 135; Conservative 93; Mismatches 206; Indels 248; Gaps 38;

QY 11 LAILINTPELLA-QAHETQSGVLETVTVGSRPRA-----TSGLLHTSTASDKIISGDT 65
Db 24 LTFSLFSLVGAKQEHHTLQKV-----TTTEQRFNPSAPLSWQSEEMRNSTSRVTISNKE 79

QY 66 LRQKA-VNLGDALDGVPGIHASQYGASAPVI--RGQTGRRIKVLNHHGEGTGM-----117
Db 80 LKXTGNLNTENALONVFGIQRDATGTGVLPIKSVRGFGGG-----NGHSNTNMLLVNGI 135

QY 118 ----ADPSP-DHAIMVDTALS-QQVEILRGPTVLLYSSGNVAGLVADVADGKIPEKMP 171
Db 136 PIYGAPYSNIELAIFPVTFSQVDRIQVIGKGTGTSVQYGNPTFGGVNIIITKEIPKEW-ENQ 194

QY 172 VSGELGL--RLSSGNL-----EK-----LTSG--GINIGLG-----198
Db 195 ABERITFWKSSNGNFVDPEKPKPLAQTILGNQMLFNTVGRTAGMLGRHVGISAGNNIN 254

QY 199 -----KNFVLHTEGLYR-----KSGDYAVPRYRN 222
Db 255 GQGFQNSPTKQVNYLL--DAVKINATNTFKAYQYQYNSYHPTGLSAQDYAVNRFIN 312

QY 223 LKXLPDS-----PRRFA-----NGQRAVLGRKRFYERTYSDRD-----QY-----260
Db 313 ER--PDNQDGRKRGFIVYQVYFDPDRKVGDFKFTYFTTHDMGRDFGFSNQYOSVYMS 370

QY 261 -----GLPAHSHEYDDCHADIIWQ-----KSLIN-----284
Db 371 SQNKILPFKKGKISATNPNGLSYSYDNTSPC-----MQFDNIRSRVNAFEPKLNLI 425

QY 285 -----KRYLYLPHLLTBE-----DVDYDNPGL-----SCGFHDDDDAHAAHNGKPW-----327
Db 426 VNTGKVKQTFNMGMRFLETDLRYRSTRKNPSMNNNGSGF-----DAGTSLNNFNNTAVY 481

QY 328 -----IDLN-----KRYELRAEWKQPPGFPEALRVHLNRNDYHDDKAGDAVENFEN 375
Db 482 ASDEIFNNGMLITITGLRYTFLNVEKQAPFKA-----CQTKTKIDRYN 528

QY 376 NOTQNAIRLHOPIGRKLSGWGV--YLGQKSSALSATSEAVKQPMLLDNKVQHSYFPG 433
Db 529 Q--WNPANVGVYKPIKELLFYPYQYSYTPPOPSNIGS-----FVGTSTDYFQIFN 577

QY 434 VEQANWNTFLEGGVVE-KQKASIRYDKALIDRENVY-----KQPLPDLGAHRQTAR 485
Db 578 V-----MEGSRYYFNQVSNFANYFVFPANNYFTGRYDKNKEPV-----NAR 620

QY 486 SFALSGNWYFTPOHKLSTASH 507
Db 621 SQGVELELYTPIRGLNFHAAY 642

RESULT 28
US-09-882-227-78
; Sequence 78, Application US/0988227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Oomen, Raymond P.
; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
```

```

; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 767
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 304
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-882-227-78

Query Match      3.0%; Score 122.5; DB 12; Length 767;
Best Local Similarity 18.0%; Pred. No. 0.027;
Matches 161; Conservative 117; Mismatches 337; Indels 279; Gaps 42;

QY 14 LLINTPLLAQAHETEQSGVLETVTVVG-----KSRPRA-TSGLLHTSTASDKIISGDTLRQ 68
Db 4 ILVSLAVLSHAHVAKTHNLERVEASGVANDKEAPLSRKSVEVRNYSRVTISNKLQTLK 63

QY 69 KA-VNLGDALDGVPGIHASQ-----YGGASAPVIRGOTGRRIKVLNHHGE 113
Db 64 SANQISEALQNVGVHIRNATGIGAVPFSVRFGGSS-----GHSNTAMVLVN-----114

QY 114 TGMADPSP--DHAIMVDTALSQOVE--ILRGPTVLLYSSGNVAGLVADVADGKIPEKMP 168
Db 115 -GIIYVAVYDVISIPFPVTFQSDRISVTKGESVRYGVNFGVGINVITKGIPTKW--172

QY 169 ENGVSGBELG--RLSSGNLEKLTSGGINIGLKNFV---LHTEGLYRKSGDYAVPRYRN 222
Db 173 ESQVSEERATFGKSENGGFFNQNSKNLDSLANMLPDTYLTGGMNKK--HFGIQAAQAN 230

QY 223 LKRLPDSRPFANGQRAVLGRKRFYRRTYSRDOYGLPAHSHEYDDCHADIIWQKSL 282
Db 231 -----WLKQO-----GFR-----YNSPTNIQNTY---MLDSLYQIND 258

QY 283 INK--RYLQLYPHLLTBEEDVDYDNPGLSCGFHDDDDAHAAHNGKPWIDLNRKRYELRAE 340
Db 259 SNKITAFFQYNYFMAD-----PGSL-----279

QY 341 WKQFPFPEALRVHLNRNDYHDEKAGDA-----VENFNNQTONARIELRHQPIORLK 394
Db 280 -----GIEAYNQNRFPQNNRPNNNKSGRAKXGWGAVYQNFQFD-----TDKIG 320

QY 395 GSWGVOYLQO-----KSSALSATSEAVKQPMLLDNKVQHSYFPG-----433
Db 321 GDTFESYGHDMSDRDFQDSNFLNVTNPKLGPVYTDQNTYPGFIFDHLRRIYNNAEFEN 380

QY 434 ---VEQANWNTFLEGGVVEKQKASIRYDKALIDRENVYKQ--PLPDLGAHRQTARSFA 488
Db 381 LNLVNTNKKVKQTFNVGMRFMTMDMYFPLDQSTCEKTD1FNGVCRMPFPFVLSKKFSSNNQ 440

QY 489 LSGN---WYFTPOHKLSTASHQERLPSTCELYA--HGKHVAINTPEVGNKHLNKRSEN 542
Db 441 LFNNYTAWL---SDKIELFDSKLVIITPGLRYTFLNKNKEPEKHDPSVNV--ITKQRON 495

QY 543 NIELAL--GYE-----GDRMOY-NLA 560
Db 496 EWSPALNIGYKPMENWIWYANYSRFFIPQHTMLGITRTNNTYQIFNIEVQRYSYKNLL 555

QY 561 LYRNR-----FGYIYAQTILNDGCPKSIEDDSEMKLYRYNQSGADFYGAEGEYFKTTPR 616
Db 556 SFNTNYFVIFAKRYYA---GGYSPQPINARSQ-----GVLELYYAPIRG 597

QY 617 YRIGVSGDYVRGRU--KNLPSLPGREDAYGNRPPTFAODONAPRVPAARLGFHLKASLTDR 675
Db 598 LQFHAVYTYIDARITSNADDIAYFTGIVNKPFDIK--GKRLPYVSPNQFIDMMYTYKHT 656

QY 676 IDANLDYVRVFAQNKL---ARYET-----RTPGHMLNLNLAN-----YRRNTRY 716
Db 676 IDANLDYVRVFAQNKL---ARYET-----RTPGHMLNLNLAN-----YRRNTRY 716
```

Db 657 TFGISSYFYRAYSSMLNQAQSQTVCLPLNPETGG---LEYCNSVGLLLPLYFVLNVQV 713
 QY 717 GWNWY-----VKADNLLNQSVYAHSSFLSDT---PQGRSPTGGVNVKP 758
 Db 714 SSVLMQSGRHKTIGSLQINNLFNNKYFYRGIGTSPTRGPAPGRSITAYLVNEF 767

RESULT 29

US-10-180-326-1

; Sequence 1, Application US/10180326

; Publication No. US20030049661A1

; GENERAL INFORMATION:

; APPLICANT: Seino, Susumu

; APPLICANT: Shibasaki, Tadao

; TITLE OF INVENTION: Protein Rim2

; FILE REFERENCE: P21573

; CURRENT APPLICATION NUMBER: US/10/180,326

; CURRENT FILING DATE: 2002-06-27

; PRIOR APPLICATION NUMBER: JP 288372/99

; PRIOR FILING DATE: 1999-10-08<160> 5

; NUMBER OF SEQ ID NOS:

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 1590

; TYPE: PRT

; ORGANISM: Mus musculus

US-10-180-326-1

Query Match 3.0%; Score 122.5; DB 15; Length 1590;

Best Local Similarity 18.6%; Pred. No. 0.09;

Matches 139; Conservative 102; Mismatches 286; Indels 221; Gaps 35;

QY 63 GDTLRKAVNLGDALDVGPI-HASQYGGGASAPVIRGQTGRRIKVLNHHGETGDMADFS 121

Db 116 GEESQOQEQKGA--PTGICHKTKEADGCHNCSCQI-----KFCARCGRVSLR 166

QY 122 PDHAIMVDTALSQVEILRGPVLLYSSGNVAGLVVDVADGKIPEKMPENGVSSELGLRLS 181

Db 167 SNKVMVYCNLCRQKEILTKSGAFWYNSG--SNTLQPDQKQVPR-----GLRNE 213

QY 182 SGNEKLTSGGINIGLGNFVLTGELYRKS-GDYAVP-----RYRNKRLPDSRRFAN 235

Db 214 EAPQEKAK-----LHQPQFQAGDLSVPVAVKEGRAHCLTR-QDTIKNGSG 260

QY 236 GQRAVLGM--RKRF-----YRTY--SDRRDQYG-----LPASHRYDDCHADI 276

Db 261 VKHQIASDMPDRKRSVSRDQNRREYQSEEREDYSQYVPSDGTWPRSPSYAD----- 315

QY 277 IWQKSLINKRYLQLYPHLLTEEDVDYDNPGLSCGFPHDDDDAHAAHNGKPWID----- 329

Db 316 -----RRSQREPQF-----YEPGHLN---YFDSNRCHRHRSKEYIVDDDEVES 356

QY 330 -----LRNKRYELR-----AEWK-QPPGFEALRVHLNNDYHHDKAGDAVENFF 374

Db 357 RDEYERQRREYQARYRSPNLARYVPKQPVBEQMRIAEVSRARHERHSDV--SLA 414

QY 375 NNQTONARIEL-----RHQPIGLKSGMGVQVILGOKSSALSATSSEAVKQ----- 418

Db 415 NAELEDRSILRMDRSPRQBSVERRAWE-----NORSISWETREARQSSYSPORTSN 470

QY 419 -----PMLDNKVQHSYFFGVEQANWDMFTLEGGRVVEKOKASIRYDKALIDRENY 469

Db 471 HSPPTPRSPITLDRPMRRADSLRKQHLLDPSSAVRKTREKMETMLRNDLSLSSQSES 530

QY 470 YKQPLPDLGAHQRTARSFALSGNWFYTPQHKLSLTASHQERLPTQELVYAHGHV---AT 526

Db 531 VRPPPPRHKKSGGK-----MRQVSLSSS--EELASTPB-YTSCDDVELESE 576

QY 527 NTFVEGNKHLNERSNIELALGYEGDRWQVNLALYRNFQNIYIAQTLNDGRGPKSIED 586

Db 577 SVSEKGDQSGKRTSE-----QGVLSDSNTSRSEKQRMYY-----GHSLEE 620

QY 587 DSEMKLVRYNOSGAD-----FYGABGEIYFKPTPRYRIGVSGDYVRGRL---KN 632
 Db 621 DLEWSEPOIKDSGVDTCSSTLTNEEHSRSDKHPVTWQPSK-----DGRLLIGRLLNKR 674
 QY 633 LPSLPGREDAYGNRPFFIAQDDQNAFPRVPAARLGFHLKASLTDRIDANLDYRVFAQNKLA 692
 Db 675 L-----KDGSVPRDSGAMGLGLKVVG-----KMTESGRLC 704
 QY 693 RYETRTPGHEMLNLGANYRNRTRYGEWN 720
 Db 705 AFITKVKKGLADTVGHLRPGDEVLEWN 732

RESULT 30

US-10-238-075-1277

; Sequence 1277, Application US/10238075

; Publication No. US20030148324A1

; GENERAL INFORMATION:

; APPLICANT: I.N.S.E.R.M.

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isol

; FILE REFERENCE: BLANDINE

; CURRENT APPLICATION NUMBER: US/10/238,075

; CURRENT FILING DATE: 2002-09-10

; PRIOR APPLICATION NUMBER: 0003145

; PRIOR FILING DATE: 2000-03-10

; NUMBER OF SEQ ID NOS: 1576

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1277

; LENGTH: 753

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-238-075-1277

Query Match 3.0%; Score 120.5; DB 12; Length 753;

Best Local Similarity 20.8%; Pred. No. 0.041;

Matches 146; Conservative 90; Mismatches 283; Indels 183; Gaps 34;

QY 30 SVGLETVTWGKSRPRATSGLLHTSTASDKIISGDTLRQKAVNLGDALDVGPGIHAS-QY 88

Db 44 SVGKTTSEALEKTGATS-----SRTTDKNLQ-----SLDATVRSMPGTVTQIDP 89

QY 89 GCGASAPVIRGQTG-RRIKVL-----NHHGETGDMADFSDPHAIWDTA 131

Db 90 GCGAISVNIKSGSGFRVNTWVDGITSFYGTSTGTTTHGSTNNMA-----GVLIDPN 143

QY 132 LSQOQVEILRGP-----VTLLYSGNV--AGLVVDVADGKIPEKMPENGVSSELGLRLSSG 183

Db 144 LLVAVDVTRGDSGSGSEGINALAGSANMRTIGVDDVI-----FNGNTYGLRSRFSVG 194

QY 184 NLEKLTSGGINIGLGNFVLTGELYRKS-GDYAVP-----RYRNKRLPDSRRFAN 239

Db 195 SNGLRSGMIALG-GKSDAFTDTGSGYMAAVSGSSVYSNFSN-----GSGINS 242

QY 240 AVLWKRKFFYRTYS-----DRDQYGLPAHSHEYDDCHADI IWOKSLINKRYLQ 290

Db 243 KEFYDKYMKQNPKSQLYKMDIRDPFNSFELSARTYENKTRDI-----TSDDYIK 296

QY 291 YPHLLTEEDVDYD-NPGLSCGFHDDDDAHAAH-----AHNGKPWIDLRLN-KRYELRAEWK 342

Db 297 YHTTFFSELIDFNVTASTSRGNQKRYDRGSLTYFTYKTSQNRSDALDINNTRSRFTVADNDL 356

QY 343 QPFPGFPEALRVHLNNDYH--HDEKAG--DAVEN--FFNNQTONARIELRHQPIGLRKGSW 397

Db 357 EFWLGSKLMRTYDRTIISAAGDPKANOESIENPFAPSGQDP--ISALYGLKVTGRIW 414

QY 398 GVOY-LGQKSSALSATSSEAVKQPMLLDNKVQHSYFFGVEQANWDMFTLEGGRVVEKOKAS 456

Db 415 EADFNLYNTRNRTITGYKPA-----CDSRV-----ICVPQGSYDIDDKEGGF----- 455

QY 457 IRYDKALIDRENYKQPLPDLGAHQRTARSFALSGNWFYTPQHKLSLTASHQERLPTQOE 516

Db 456 -----NPSVQLSAQ-----VTPWLPQFFIGYSKMRAPNIQE 486

QY 517 LYAHGHVATNTFEVGNKHLNKRNS-----NIELA-LGYEGDRWQYNLALYRNRFGNY 569
Db 487 MFFNSGGASM-----NPLKPERAETWQAGFNIDTRDLLVEQDALRFPKALAYRSIQNY 541
QY 570 IYAAQ-----TLND--GRGPKSIEDDSEMKLVRYNOSGADFYCAGBEIYFKPTPR 616
Db 542 IYSESVLVCGRKSLPEVINGWEGISDESDNMYIYNSASDVI-AGK---FELEMD 597
QY 617 YRIGVSGDYVRGLKNLPSLPGRDAYGNRPPIAODQNAEP 658
Db 598 YDAG----FAFGRLSFSQOQDPTSIASHTFGAGDITELPR 635

RESULT 31
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR FILING DATE: 1998-12-11
; PRIOR FILING DATE: 1998-01-16
; PRIOR FILING DATE: 1998-01-16
; PRIOR FILING DATE: 1995-06-07
; PRIOR FILING DATE: 1998-01-20
; PRIOR FILING DATE: 1995-06-07
; PRIOR FILING DATE: 1995-06-07
; PRIOR FILING DATE: 1998-01-16
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 2.9%; Score 118.5; DB 9; Length 1475;
Best Local Similarity 18.7%; Pred. No. 0.19; Indels 227; Gaps 35;
Matches 125; Conservative 85; Mismatches 230;
QY 87 QYGGGASAPVIRGQTRRIKVLN--HHGETGDMADFPDHAIMV---DTALSQQVEILRG 141
Db 894 RYDLGISKPNKYGTADDLVKAIRKALHSKGIKWADWVPDQMYAFPEKEVVTATRVDKYGT 953
QY 142 PV-----TLIVSGNVAG-----LVDVADGKTP-----KMPENGYSSELGLRL 180
Db 954 PVAGSQIKNTLYVVDGSKSGKQOAKYGGAFLELOAKYPELPARKQISTGVPMDSVKI 1013
QY 181 SSGNLEKLTSGGINIGLQKPFVLHTEGLYRKSGDYAVPRYRL---KRLPDGPRRPAQC 237
Db 1014 QKWS-AKVFNGTNILRGAGVVLK-----DQANTYFNISDNKEINFLPKTLNOD 1063
QY 238 HRAVLGWRKRY-----RTYSDRDRDQVGLPAHSHEYDDCHADIIWQSLINKRY 287
Db 1064 SQVGFSDYDGKGVYVYSTSGYQAKNTFISGDKW-----YFONGNMTVGTGAQSGVNY 1117
QY 288 -----LQYPLHLLTEEDVDYDNPGLSCGPHDDDDAHAAHNGKPMWIDLNKRYE----- 336
Db 1118 YFLSNGLOLRALKKNEDGT-----AYGND-----GRYENGYQ 1154
QY 337 -LRAEWKOPFGFEALRVHLNRND-----VHDE-----KAGDAVENP 373
Db 1155 FMSGVWR-----HFNNGMSVGLTVIDGQVQYFDEMGMGYQAKGFVTTADGKIRY 1203

QY 374 FNNQTONARIELRHQPIGRLKSGWGVQYLGQSSMLS-----ATSEAVKQPM 420
Db 1204 FDKQSGNM---YRNRFIENEKRW--LYLGEDGAAVTSQTINGOHLIFRANGVOVKGEF 1258
QY 421 LLDN--KVQHYS-----FFGVEQANWDF-----TLEG-----QVR 449
Db 1259 VTDHGRISYDGNSGDQIRNRFVRNAQGWFFDNNGYAVTGTARTINGQLLYFRANGVQ 1318
QY 450 VEKQKASIRYDICALIDRENYKQPLDLCAGHPTARSA--LSGNWYFTPOHKLISLTASH 507
Db 1319 VKGEFVTDY-----GRISY-----DGNSGDOIIRNRFVRNAQGWFFDNNGYAVTGA- 1367
QY 508 QERLPSTQELIYAHGKHVATNTFEVGNKHLNKRNSNIELALGYEGDRWQYNLALYRNR- 566
Db 1368 --RTINGQHLIFRANGVQVKGEFVTDH-----GRISYDGNSGDQ-----IRNRFV 1412
QY 567 -----GNVYTAQTLNDGRGPKSIEDDSEMKLVRYNOSGADFYCAGBEIYFKPTPRRIGVS 622
Db 1413 RNAQGWFFD--NNG-----YAVTGARTINGQ-HLYFRAN---GVQVK 1450
QY 623 GDYVRGR 629
Db 1451 GEFVTDH 1457

RESULT 32
US-09-815-242-5247
; Sequence 5247, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5247
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5247

Query Match 2.9%; Score 117.5; DB 9; Length 422;
Best Local Similarity 18.3%; Pred. No. 0.031;
Matches 91; Conservative 69; Mismatches 149; Indels 189; Gaps 21;
QY 61 ISGDTLRQKAVNLGDALDVGFGHISYGGGASAPVIRGOT--GRIKVLNHHGTTGDMA 118
Db 27 VDTDLRLKLABQVDDI-----VFISGTNGKTTSLNLTGKANNIQI--HNNEGANNA 79

QY 119 DFGPDHAIWDTALSOQVEILRGVPTLLYSSGNVAGLVADVADGKIPKMPENGVS----- 173
 DB 80 AGITSAFIMQST-----PKTKI-----AVIEIDEGSIPLVLKEVTGPMVFT 121
 QY 174 -----GELGLRSLSSNLEKLTSGGINIGLGNFVLHTEGLYRKSGDVAVRYRNL 223
 DB 122 NFRDQMDRGEIDIMWNN-IAETISNKGKIKLL-----NADDFVSLR----- 164
 QY 224 KRLPDSPPRFANGQRAVLGRKFRYRTYSDRRDQYGLPAHSHEYDDCHADIIWQKSLI 283
 DB 165 -----KIASDTIVVYGMKAHAHEP-----QSTWN 189
 QY 284 NKRYLQYPLHLLTEEDVDNPG-----LSCGFHDDDDAHAAHNGKPDWIDLRNKRVELRAE 340
 DB 190 ESRYCPMCGKLLQVYIHYNQIGHYHCQCGKKE-----QAKYEISF 232
 QY 341 WKOPFGFEALRYHLNRNDYHDEK-AGD-----AVEN----- 372
 DB 233 DVAPP-----LHLNINDEKYMKIAGDFNAYNALTAVTLRELGLNEQAIKNGFETY 285
 QY 373 -----FFNQTNARIELRHOPIG-RLKSGWQVYLQKSSALSATSEAVKQPMILDN 424
 DB 286 SDNGRMQYFKKERKEMINLAKNPAGNALSUSVGEQLEGEKVVIS-----LNDN 335
 QY 425 KVQHSYFPGVEQANWDNFT-----LEGGVRVEKOKASIRYDKA-----LIDRENYKOP 473
 DB 336 AADGRDTSWIYDADPEKLSKQTEAIIIVTGTAEELQLRLKLAEEVPIIVERDIY----- 391
 QY 474 LPDLGAHQRTARSFALSG 491
 DB 392 -----KATAKTMDYKG 402

RESULT 33

US-09-815-242-5456
 ; Sequence 5456, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5456
 ; LENGTH: 978
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-5456

Query Match 2.9%; Score 117; DB 9; Length 978;
 Best Local Similarity 18.9%; Pred. No. 0.14;
 Matches 162; Conservative 103; Mismatches 277; Indels 316; Gaps 44;
 QY 19 PLLAAQHETQSQVLETV---TVVGKSRPRATSGLLHTSTASDKIISGDTLQKAVNLCD 75
 DB 73 PKAVQAPQTAQANLETIVKEEVKERAKPQ---VKETTPQDN---SGD---QKQVDL--- 121
 QY 76 ALDGVPGIHASQYGGGASAPVIRGQYGRRIKVLNHHGTEGDMADFPDHAHMDALSOQ 135
 DB 122 -----TPKMATQ-----NQAAET-----Q 135
 QY 136 VEILRGVPTLLYSSGNVAGLVADVADGK-IPKMPENGVSSELGLRLSSGNLEKLTSGGIN 194
 DB 136 VEVAQ-PRTVSESNPRVRSADVVKEASDAKVEITDVTSKVTVESGSEIAPQGNKYE 194
 QY 195 IGLGNKFLV-----HTEGLYRKSGD-----YAVPRYRNKRLPDSPPRFANGQ 237
 DB 195 PHAQRVVLKYKLFKEKGLHK--GDYDFDTLSNNVNTYGVSTARKVPEIKNGSVVVMATQ 252
 QY 238 HRAVLGWRKFRYRTYSDRDQYGLPAHSHEYDDCHADIIWQKSLINKRYLQYPLHLLTE 297
 DB 253 ---LLGNK--IRYTFDYID-YKVNVT-----DLEINLFIDPKTVQSGQQTITSLND 302
 QY 298 ED-----VDYDNPLSGCFPHDDDDAHAAHNGKPMIDLRNKYE----- 336
 DB 303 KETKNTLPIEY-NPGVS-----NSYANVNGSIETFDKGNRFTTHVAVIKPONGHKSDS 354
 QY 337 -----LRAEWKOP-FPGFEALRYHLNRNDYHDEKAGDAVENFFNQTNARIEL 385
 DB 355 VSIQTTLTQSGKADGKAPTIVKYEVLK---DANELQSVYANVSDSMFADVTTEWMDKL 411
 QY 386 RHQ-----PIGLKSGWQVYLQKSSALSATSEAVKQPMILDNKVQHSYFPGVE--- 435
 DB 412 KVENNGNVKLDIEKLSYVIHYDGE---YLSGSDQVNFRTMFGYPEQYKYVYTHLGY 468
 QY 436 QANWDN-----FTLEGGVRVEKOKASIRYDKALIDR-ENY 469
 DB 469 KLTWDNGLVLYSNKAKGDTNGTITESNNMTFDEEYGTGV---ITQVDKKNLVTTVEE 524
 QY 470 YKQPLPLGAHR-----QTARFALSAGNMYFTFQ---HKLSTASHQE 509
 DB 525 YDSSTLDDIYHTAIDGEGYVDGYETIETDSSAIDYHTAVDSEAGHVGGVYTESSEE 584
 QY 510 RLP-----STOELYAH-----GKIVAT--NTFEVGNKHL----- 536
 DB 585 SNPIDFEESTHENSKEHADVVVEEDTNPGGQVTTESNLVFEFDESIGIVTGAVSDHT 644
 QY 537 ---NKE---RSNNIEL-----ALGYEGDRWQYNLAL 561
 DB 645 TVEDTKEVTTESNLIELVDELPEHGOAQGPBEITENNHHISHSLGTE----- 694
 QY 562 YRNRFNVIYAQTLDN-----GRGPKSIEDSEMKLVRYNQSG-----AD 601
 DB 695 ---NGHGNVGVIEEIEENSHVDIKSELGYEGGQNSQSFEEEDPKPKYEQGQNVIVID 752
 QY 602 FYGA-----EGEIVF-----KETPRYRIGVSGDYVRGLKNLPSLPGREDAYGNRPF 649
 DB 753 FDSVPQIQGQNGNQSFEEEDTEKDKPKYEQG--GNIIDIDFSDVPQIHG-----FNKHEI 806
 QY 650 AQDDONAPRVPAAALGFH 667
 DB 807 IEDTNKDK-PNYQFGGH 823

RESULT 34

US-09-815-242-12686
 ; Sequence 12686, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.

```

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12686
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12686

Query Match      2.9%; Score 117; DB 9; Length 1001;
Best Local Similarity 18.9%; Pred. No. 0.14;
Matches 162; Conservative 103; Mismatches 277; Indels 316; Gaps 44;

Qy 19 PLAAQHEHQSGVLETV---TVGKSRBRATSGLLHTSTASDKIISGDTLRQKAVNLGD 75
Db 96 KFAVQAPQAPLANLEKEEVKEAKFQ---VKETIQPDN---SGD---QKQVDL-- 144
Qy 76 ALDGVFGIHASQYGGGASAPVIRGQTRRIKVLNHHGETGMADFPSPHAINMVDLTALSCQ 135
Db 145 ---LRLGPGVTLLYSSGNVAGLVADVAGK-IPEKMPENGVSGLGLRLSSGNLEKLTSGGIN 194
Qy 136 VEIILRGPGVTLLYSSGNVAGLVADVAGK-IPEKMPENGVSGLGLRLSSGNLEKLTSGGIN 194
Db 159 VEAQ-Q-PRIVSESNPRVTSADVVVEAKEADAKVETGTDTVTSKVTVESGSIAPQGNKVE 217
Qy 195 IGLGKNFVL-----HTEGLYRKSGD-----YAVPRYRLKRLPDSPPRFANGQ 237
Db 218 PHAQORVLKYKLKFEKGLHK--GDYFDFTLSSNNVTVGYSTARKVPEIKNGSVVMATGQ 275
Qy 238 HRAVLGWRKFRYRTYSDRDQYGLPAHSHVEDDCHADIIWQKSLNKRYLQLYPHLLTE 297
Db 276 ---LLGNGK--IRYTFIDYID-YKVNVT---DLBINIFIDPKTVQSGNGQOTITSLND 325
Qy 298 ED-----VDYDNPGLSCGPHDDDDAHAAHNGKPMIDLNRKYE----- 336
Db 326 KETKNTLPIEY-NFGVS-----NSYANVNGSIETFDKGNRFTHVAYIKPQNGHKSDS 377
Qy 337 -----LRAEWKQP-FPGFEALRVHLNRNDYHHDKAGDAVENFENNQTQVAREL 385
Db 378 VSTIGTLTGSKADGKAPTQKVVEVLK---DANELFQSVYANVSDSSMPFADVTEEMKOKL 434
Qy 386 RHQ-----PIGLKGSWGVQLGQKSSALSATSBAVKQPMLLDMKVQHYFFGVE--- 435
Db 435 KVENNGVYKLDIEKLEKSYVIHVDGE---YLSGSDQVNFRTMFGYGEQYKYYTHLGY 491
Qy 436 QANVDN-----FLEGVRVEKOKASIRTDKALIDR-ENY 469
Db 492 KLTWDNGLVLYSNKAKGDGTNGTITENNNMTFDERYGTGV---ITGOYDKNLVTTVEBE 547

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12686
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12686

Query Match      2.9%; Score 115.5; DB 9; Length 981;
Best Local Similarity 19.3%; Pred. No. 0.19;
Matches 141; Conservative 129; Mismatches 265; Indels 197; Gaps 38;

Qy 86 SQYGGGASAPVIRGQTRRIKVLNHHGET-GMADFPSPHAINMVDLTALSCQVEILRGPT 144
Db 69 NOYGGKL---VLIDDDGLEIEVERIKGSAQGDVKVYLPNGAVRDDAWLQKLNYSKKTY 125

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12211
; LENGTH: 981
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12211

Query Match      2.9%; Score 115.5; DB 9; Length 981;
Best Local Similarity 19.3%; Pred. No. 0.19;
Matches 141; Conservative 129; Mismatches 265; Indels 197; Gaps 38;

Qy 86 SQYGGGASAPVIRGQTRRIKVLNHHGET-GMADFPSPHAINMVDLTALSCQVEILRGPT 144
Db 69 NOYGGKL---VLIDDDGLEIEVERIKGSAQGDVKVYLPNGAVRDDAWLQKLNYSKKTY 125

```

QY 145 LLYSSGNVAGLVADGKIPKMPENGVSGBELGLRLSSGNLEKLTSGGINIGLGNFV 202
 Db 126 QGIFSDVLGLQDI-HRNLNEKQLQYLLQAGALSTGTEFTSMREVNR 172
 QY 203 LHTEGLYKSG--DYAVPRYRNK-----RLPD-----SPREFANGQHR 239
 Db 173 -KKDELYKSGKMPIINQIQTEQLQLESQIREEEKLETHVRLVDORDKSRRLNLRN 231
 QY 240 AVLGMWRFRYRTYSDRRDQYGLPAHSHEYDDCHADIIMQKSLINKRYLQYLPHTLTED 299
 Db 232 L-----NQLSKMHEEKQKVALHDSQEW-----KSL--BQQLNISPITPEKG 273
 QY 300 VD-YDNFGLSCGFHDDDAHAHNGKPMIDLRKRY-ELRAEMKQFPF----- 346
 Db 274 VDRYEK-----ARAHKQSLERDGLRNERLAQKEBATOEPVKOSDDIDAFIS 321
 QY 347 -----GFE-----ALRVHLNRNDYHDEKAGDAVENFNNQ 379
 Db 322 LNOQENEIKNKEPETAIEKDIAKQKQKDELQANIGWSETHDVSSEAMKSYVSEQIK 381
 QY 380 NARIELRHQPIGLKSGWQYLGQKSSALSATSEAVKQPMLLD---NKVQVYSPFGEQ 436
 Db 382 NKQQAAY--IKQLERLEENKI--EDNAVHSELDSVEEKIVPEETFEKKKYSQQVIEL 437
 QY 437 ANWDFN--TLEGGVVRVKKASIRYDKALIDRENYKQPLDILGAHROTARSFALSQWY 494
 Db 438 NEKENLYSKLERFEIEQEKQKQK--QKLL--RTTFILLTLVGIGL--TAFSP-IGNNML 490
 QY 495 F-----TPQKLSLTASHOERLPSTOELYANGKHVATNTFVGNKH-----LNKERSNNI 544
 Db 491 FGIIFAVLTVFVIGIMSKSKEYDYSEAL-----TDEIBEIKQALIDENYDLDF 542
 QY 545 ELALGYE-GDRWQYNL-----ALYRNRFNGYIYAQTLDNDRGPKSIEDDSMKLVRYNQSG 599
 Db 543 DLDEQYRDRHWQALKNKQKILEKR--QYIEGR-LNDKAG-----RHDELQ 586
 QY 600 ADFVGAEEIYFKTPRYRIGVSGDYVRGLKNLPSPGPRE-----DAYGNRPFIADQ--- 652
 Db 587 STVENKDELILYSK-----ISNDLIVDSISTWANTKALDQIHSDINQOQQLVOELDT 640
 QY 653 -DONAPRVFARLGFHLKASLTDR-----DANLDYRVFPAQNKLARYETRTPHHM 703
 Db 641 FYNHAEAVTKSQVYFNKLSLFDVQWLGKSAEDTN-EKWRINAEN-----IKLVNTEL 693
 QY 704 LNLGNRYRNTR 715
 Db 694 NHLNQAQLENNK 705

RESULT 36

US-10-260-877-62
 ; Sequence 62, Application US/10260877
 ; Publication No. US20030021813A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Abbott Laboratories
 ; APPLICANT: Chovan, Linda E.
 ; APPLICANT: Hessler, Paul E.
 ; APPLICANT: Reich, Karl A.
 ; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
 ; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
 ; TITLE OF INVENTION: 'ESSENTIAL GENES'
 ; FILE REFERENCE: 6565.US.P1
 ; CURRENT APPLICATION NUMBER: US/10/260,877
 ; CURRENT FILING DATE: 2002-09-30
 ; PRIOR APPLICATION NUMBER: US/09/649,145
 ; PRIOR FILING DATE: 2000-08-25
 ; NUMBER OF SEQ ID NOS: 137
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 62
 ; LENGTH: 631
 ; TYPE: PRT
 ; ORGANISM: H. influenzae
 ; US-10-260-877-62

Query Match 2.8%; Score 112; DB 15; Length 631;
 Best Local Similarity 18.7%; Pred. No. 0.2;
 Matches 128; Conservative 76; Mismatches 240; Indels 240; Gaps 35;
 QY 154 GLVDVADGKIPKMPENGVSGBELGLRLSSGNLEKLTSGGINIGLGNFV 201
 Db 80 GTITISONGI--GWTREQVIDHLGTAKSGTKEFLTALGQDQAKNSQLIQPGVGYSAF 137
 QY 202 VLHTEGLYRK-----SGDYAVPRYRNKRLPDSPPRFANGQHRVAVLGM 244
 Db 138 IVADKVTVKTRAAEEADKAVLWESAGEGEYSVADTEKKGRGTDVILHLREDEKEFLNEW 197
 QY 245 RKRFRYRTYSDRRDQYGLPAH--SHEYDD---C-----HADIIWOKS---LINKRY 287
 Db 198 RLRIIIGKYS---DHIGLPVEMLTKEYDDGCKGCKWEKINKSDALWTRKNDVDEEY 254
 QY 288 LQLYPHLLTEEDVDYDNPGLSCGFHDDDAHAHNN-----GKPMIDLR 331
 Db 255 KAFYKHL-----SHDFVDPVTWAHNNKVEGNQAYTSLLYVPAKAPW-DLF 297
 QY 332 NKRYELRAEWKQFPFGEALRVHLNRNDYHDEKAGDAVENFNNQTONAR--IELRHQP 389
 Db 298 NREHK-----HGLKLYVQVRFIMDD-----AEQFIPNYLRFMRGLIDSNLDLP 339
 QY 390 IGRLLKSGWQYLGQKSSALSATSEAVKQPMLLDNKVQHYSEFGVEQANWDFLEGQVR 449
 Db 340 LN-----VSREILQDNKI-----TAALR 357
 QY 450 VEKQASIRYDKALI--DRENY---YKQ-----PLPDLAGHROTARSFALSQWYFT 496
 Db 358 KALTKRSQMLKLEKADDAEKYLOFWKEFGLVLKGEPAEDF-ANKETVAKL----- 407
 QY 497 PQHKLSTASHOERLPSTQEL-----YAHGK-----HVAINTFEV--GNKHLNERSNNI 544
 Db 408 ---LRFASTNDSEGTVSLEDYILRMKSGQKAIYITADSYAAKNSPHLELFNKKG 463
 QY 545 E-LALGYEGDRWQYNLALYRNRFNGYIYAQTLDNDRGPKSIEDDSMKLVRYNQSGADFY 603
 Db 464 EYLLLSDRIDRWMLLS---YLTEF-----DGKQLOSI-TKADLDL-----GDLA 502
 QY 604 GAEGEIEYFKTPRYRIGVSGDYVRGLKNLPSPGPREDAYGNRPFIADQDQAPRVPAAR 663
 Db 503 DKSE-----TQKQDEAFSGFIE-RVKNL-----LGERVKTVRLTHNLTDIPA-- 545
 QY 664 LGFHLKASLTDRIDANLDYRVFPAQNKLARYETR-----TFGHMLNGLNRYRNTRYGE 718
 Db 546 -----VVSTDNQWTTQMAKLFRAAGQPVPEVYTFELNPEHLVKKVADIADETEFAD 599
 QY 719 WNVYVKADNLLNQSVYAHSSFLSD 742
 Db 600 W-----VELLLEQANLAERGSLEN 618

RESULT 37

US-09-738-626-6493
 ; Sequence 6493, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18

RESULT 38
US-10-055-364-44
; Sequence 44, Application US/10055364
; Publication No. US2002015433A1
; GENERAL INFORMATION:
; APPLICANT: Patience, Clive

Query Match	2.7%;	Score 109.5;	DB 14;	Length 818;	
Best Local Similarity	17.7%;	Pred. No. 0.54;			
Matches 110;	Conservative 96;	Mismatches 221;	Indels 195;	Gaps 29;	
QY	264	AHSEYDDCHADIIWOKSLINKRYLQY----	PHLLTEEDVDYDNPGLSCGFHDD	315	
DB	123	AITNQYTSSFAMPLEARLUDYNY-ECYNGIQVTENGHLTYVDVDGYNESVRLVPADGL	181		
QY	316	DAHAHANGXP--WIDLNRKY-----	ELPAEWKQPPGPEALRVHLENRDY	360	
DB	182	TSSIRRYHSPELYVTPRNLMSYTRTTVNCVEIDMTARSHKPFYEVTT	231		
QY	361	HHDEKAGDAVEN--FFNNOTONARIELRHQ----	PIGRLKSGMWQVYLQKS----	408	
DB	232	-----ASGSIETSPFFYNASRRVPVOVLVYSVTDYGVGLSGENVTRFFATLNDFFISW	287		
QY	409	LSATSEAVKQPMLLDNKKVQHYSFPGVEQANWDFNTEGGRVREKOKA----	SIRYD	460	
DB	288	KAAATENSYSCLPLV-----	WKGFF--PSAIQTCKEKS YHF IADAVTASFT	329	
QY	461	KALDRENNYK-----	QPLPDLCAHQROTARSAFSLGNWY--FTQHK	500	
DB	330	TPLTDETSYNTTYQCAWODIEGIEQKRPDPVSKTHARNGSVQIYKTSGNLYVVMQPLVQ	389		
QY	501	LSLTASHQERLPSTQELYAHGKHVATNTF-----	EVGNKHNLKERSNNIELALGY	550	
DB	390	LDLLAAHAKTINSTDN--STSPPTAPNTTSTSSRRKRDRDGTGNTATNNSNNSSM----	443		
QY	551	EGDRQWYNALYNRFGNIYIYAQTLNDRGPKSIEDDSEM-----	KLVRYNQSGADP	602	
DB	444	-----EENLATSQVQFAYDQLRKSIN-----	RVLSQLRVMQCNQYRASLMMWYELSKINP	493	
QY	603	YGABGEYTFKPTPRYRIG-----	VSGD-----YVRGLKNLPSLPREDAYGNRPPIA	650	
DB	494	TSVMASLYGRPVSAKLVDGVVQISDCITVDQESVFHNR-----	VPGSKDLYTRPVWG	549	
QY	651	QDDQNASRPVPAARLGFH-----	LKASUTDRIDANLOYRVPFQON--KLARYE-----	TR	697
DB	550	FKFINGSELFVGOLGARNEILLSTNLVEVCOHSECHYFQGGNHYYKKNYEVYSTMTLTD	609		
QY	698	TPG--HMLNLGANYRN-----	TRYGBWNYVYKADNL-----	728	
DB	610	VPTLHTMITNLISVENVDQVQLYSQEKKLSNVFDIETMFREYNTYTO--NLKGLRK	667		
QY	729	-LNOGSVY-----	AHSSFLSDTPQ	745	
DB	668	DLDDSIHDGRDSFTQFLGLDQV	689		

RESULT 39
US-10-022-461-5
; Sequence 5, Application US/10022461
; Publication No. US20020142420A1
; GENERAL INFORMATION:
; APPLICANT: Haake, David A.

Fri Nov 21 10:34:26 2003

us-09-936-377-2.rapb

Page 27

Search completed: November 14, 2003, 11:05:46
Job time : 47 secs